Titre: Title:	The assignment of trips to a road network for the analysis of equitable transport
Auteur: Author:	Timothy Spurr
Date:	2010
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Spurr, T. (2010). The assignment of trips to a road network for the analysis of equitable transport [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/473/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/473/
Directeurs de recherche: Advisors:	Robert Chapleau
Programme: Program:	Génie civil

UNIVERSITÉ DE MONTRÉAL

THE ASSIGNMENT OF TRIPS TO A ROAD NETWORK FOR THE ANALYSIS OF EQUITABLE TRANSPORT

TIMOTHY SPURR

DÉPARTEMENT DES GÉNIES CIVIL, GÉOLOGIQUE ET DES MINES

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE CIVIL)

JANVIER 2011

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

THE ASSIGNMENT OF TRIPS TO A ROAD NETWORK FOR THE ANALYSIS OF EQUITABLE TRANSPORT

présentée par : SPURR Timothy

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. SAUNIER Nicolas, Ph.D., président

M. CHAPLEAU Robert, Ph.D., membre et directeur de recherche

Mme MORENCY Catherine, Ph.D., membre

M. ALECSANDRU Ciprian, Ph.D., membre

DEDICATION

To my family…

ACKNOWLEDGEMENTS

Thank you first and foremost to Prof. Robert Chapleau – the supervisor of this research project – for his encouragement, generosity, passion and integrity. The years spent under his guidance constitute a powerful and profoundly influential experience. I gratefully acknowledge the collaboration of the members of Groupe MADITUC, especially Bruno Allard, Daniel Piché and Guillaume Bisaillon, for their technical wisdom. Thanks as well to Francine Leduc of the City of Montreal, who gave me the opportunity to participate in the practice of planning major urban road infrastructure while writing this thesis. Thank you as well to the National Science and Engineering Research Council of Canada whose generous financial support made this research possible.

During my time at École Polytechnique I have had the pleasure of knowing many graduate students in transportation. I thank them for their companionship as well as their intellectual and moral support. I must also acknowledge the contribution of my family and friends to the completion of this time- and energy-intensive project. Finally, to Andréane, for helping me with the French language, but especially for your love and support, thank you.

RÉSUMÉ

Depuis plusieurs années, le transport durable et le rôle de l'automobile dans les grands centres urbains suscitent beaucoup d'intérêt. La congestion routière, la pollution, les gaz à effet de serre, la consommation énergétique et l'étalement urbain ont motivé le développement de nombreuses stratégies qui visent à modifier le comportement des voyageurs, surtout à l'intérieur des villes. En même temps, les discussions concernant la meilleure façon de financer les grandes infrastructures en transport cherchent à trouver un compromis entre les principes d'efficacité et les principes d'équité. Cette thèse a comme objectif de contribuer, d'une manière modeste, à ces discussions en clarifiant les enjeux entourant le « problème » du transport routier.

Le projet de recherche décrit dans cette thèse est fondé sur deux concepts déjà bien documentés : une analyse des systèmes de transport urbain totalement désagrégée basée sur l'information, de même que sur l'équité géopolitique en transport. Dans le cas actuel, l'analyse désagrégée est appliquée à un sous-échantillon de l'enquête origine-destination de la grande région de Montréal effectué en 2003. Le concept d'équité géopolitique est appliqué à l'étude des 15 ponts qui relient la ville de Montréal au réseau routier nord-américain. Ces grandes infrastructures jouent un rôle essentiel de redistribution parmi les nombreuses municipalités qui font partie de la grande région métropolitaine. Outre la redistribution des personnes et des marchandises, les ponts permettent également la redistribution des coûts externes du transport, spécifiquement la congestion routière, la pollution et le bruit.

L'enquête origine-destination a demandé à chaque conducteur d'indiquer lequel des ponts ils ont utilisé pour compléter leurs déplacements. Dans cette recherche, les réponses ont été soigneusement examinées afin de construire un sous-échantillon valide des déclarations de pont pour la période de pointe du matin d'une journée moyenne de la semaine. Le sous-échantillon validé était composé de 8 583 observations. En même temps, un modèle détaillé du réseau routier, comprenant plus de 100,000 liens et 70,000 nœuds, a été construit. Les liens étaient classifiés par fonction et par juridiction (municipale, provinciale ou fédérale). Puisque ces simulations étaient basées sur l'approche totalement désagrégée, elles ne comportent aucun système de zone, aucun centroïde et aucune matrice origine-destination.

Deux types de modèles ont été construits. Le modèle de validation affectait chaque déplacement du sous-échantillon de l'enquête au plus-court chemin passant sur le pont déclaré. D'autre part,

le modèle de simulation visait à prédire le choix de pont selon divers heuristiques fournis par la théorie du choix discret et par TRANSIMS, une plate-forme de modélisation basée sur les activités. Puis, l'efficacité et la pertinence de chaque modèle ont été discutées. Tous les modèles de simulation reproduisaient correctement 75% des choix de pont observés dans l'enquête origine-destination. Une comparaison des itinéraires générés par les modèles de simulation et de validation a facilité une analyse détaillée des erreurs de prédiction des modèles de simulation. Le modèle de validation a également été utilisé pour la mesure des effets redistributifs des coûts externes de transport parmi différentes population à l'intérieur de la grande région de Montréal.

Les résultats de ce travail ont mené à quelques contributions originales aux domaines de la simulation des réseaux routiers et à la discussion du financement équitable des grandes infrastructures routières. D'abord, une méthode qui utilise une matrice de confusion pour identifier et corriger des biais dans un modèle de choix de pont. Deuxièmement, un examen détaillé de phénomène d'indifférence des conducteurs face à de multiples options équivalentes, le distinguant de deux autres types d'erreur de prédiction : la déviance et l'erreur grossière. Finalemement, une des premières applications du logiciel TRANSIMS à un échantillon de déplacements observés, qui sont normalement synthétisés à partir d'autres informations.

En ce qui concerne l'équité géopolitique en transport, cette thèse examine non seulement les effets redistributifs des infrastructures routières parmi différentes municipalités, mais aussi les effets de distorsion amenés par les gouvernements supérieurs. Cette analyse est rendue possible grâce à la classification des liens du réseau par juridiction. En plus, une comparaison du modèle de simulation avec des indicateurs agrégés de consommation et de l'offre des services de transport routier générés par le modèle de validation, a seulement révélé des différences mineures. Ceci suggère qu'un modèle de simulation (prédictif) serait applicable à l'analyse de l'équité de divers projets ou politiques hypothétiques. Le résultat final de ces explorations est la formulation d'un système de péréquation entre les municipalités de la grande région de Montréal pour le financement plus équitable des grandes infrastructures routières comme les ponts. Ce système, qui vise la parité entre les différentes municipalités, serait basé sur les patrons de consommation et d'offre des services de transport observés. Enfin, cette thèse permet de disséquer les notions de coûts liés à la congestion et l'utilisation de méthodes de simulation conventionnelles permettant de les estimer. Cet exercice permet alors une discussion éclairée sur le transport équitable et l'imputation des coûts externes du transport routier.

ABSTRACT

The past several years have witnessed a growing interest in sustainable urban transportation and a re-evaluation of the role of the automobile in large urban areas. Traffic congestion, air pollution, greenhouse gas emissions, energy consumption and urban sprawl are all topics that have stimulated the development of various strategies that aim to change the way people travel, especially within cities. At the same time, the issues surrounding the best way to finance major transportation infrastructure are framed in a debate about efficiency versus equity. This thesis proposes to contribute to these discussions by clarifying, to a modest degree, the "problem" of urban automobile travel.

The research described in this dissertation is founded on two already-documented concepts: the totally disaggregate information-based approach to urban transportation simulation and geopolitical equity. Following the precepts of the former, this research uses data contained within the 2003 Montreal travel survey. With regard to the latter, the research subjects are the 15 bridges that connect the island-city of Montreal to the mainland. These infrastructure elements play a vital role in the redistribution, among the dozens of municipalities within the urban region, of people and goods and of the external costs of travel, particularly traffic congestion, air pollution and noise.

The travel survey asked automobile drivers to indicate which major bridge they used over the course of their trip. Their responses were meticulously examined to construct a valid sub-sample of declarations describing bridge usage patterns during the a.m. peak period of a typical average weekday. The final sub-sample contained 8,583 observations. Meanwhile, a model road network of the Greater Montreal Area was constructed. This complete network contained over 100,000 links and 70,000 nodes. The links are categorized by functional class and by jurisdiction (municipal, provincial or federal). Since all the simulations are based on the totally disaggregate approach, there is no zone system, no centroids and no origin-destination matrix.

Two types of models were estimated. The first type – called the validation model – assigns each trip in the survey sub-sample to the shortest path containing the declared bridge. The second type – the simulation models – attempts to predict the choice of bridge using various heuristics provided by discrete choice theory and by the activity-based modelling platform TRANSIMS. The usefulness and relevance of each of the different models are discussed. All the simulation

models correctly reproduce around 75% of the observed bridge choices. Comparisons of the itineraries generated by the validation and simulation models permits a detailed analysis of model prediction errors. The validation model is also used to estimate how the costs and benefits of travel associated with the major bridges are redistributed among different population groups within the Greater Montreal Area.

The results of this work yield several original contributions to the discipline of automobile simulation and to the discussion of equitable financing of road infrastructure. First, this research describes a method for using a confusion matrix to identify and in some cases to correct biases in a model of road facility choice. Second, this research examines in detail the phenomenon of driver indifference toward multiple equivalent alternatives, and distinguishes indifference from other types of model error, specifically deviance and gross error. Third, this project represents one of the first applications of real travel demand data to a state-of-the art disaggregate traffic simulation platform (TRANSIMS). Most users of TRANSIMS must synthesize their travel demand data from aggregate information.

With regard to geopolitical equity, this thesis examines not only the redistribution effects among different municipalities, but also the distorting effects caused by the provision of major transport infrastructure by the superior levels of government. This analysis is made possible by the classification of network links by jurisdiction. In addition, a comparison of the aggregate indicators of consumption and supply by the validation model with those generated by a simulation model revealed only small differences, demonstrating that a simulation model could be applied to evaluate the equity dimension of hypothetical transportation policies. The result of these investigations is the formulation of an equalization mechanism for the realization of parity between the numerous municipalities of the Greater Montreal Area based on their respective consumption and supply patterns. It is proposed that such a system would provide an equitable basis for the financing of major road infrastructure like the major bridges. Finally, this research dissects the notion of the costs of congestion and the use of conventional simulation methods to estimate them. Such an exercise is essential to an enlightened discussion about the costs and benefits of automobile travel.

CONDENSÉ EN FRANÇAIS

Les ingénieurs et planificateurs en transport sont très souvent impliqués non seulement dans la conception et l'optimisation des systèmes, mais aussi dans l'évaluation des impacts de ces systèmes sur le grand public, les populations particulières et l'environnement naturel. Ces responsabilités supplémentaires exigent le développement de nouvelles méthodes qui permettent de mieux comprendre les systèmes humains en général et le système de transport en particulier.

Pour ce faire, il sera question du processus décisionnel qu'ont les conducteurs lorsqu'ils choisissent les grandes infrastructures à emprunter pour un déplacement. À ces fins, le développement d'un modèle de circulation routière permettra la quantification des coûts et bénéfices liés au transport, de même que leur distribution parmi les différentes populations. Celle-ci est plus particulièrement évaluée selon le principe d'équité, qui est un enjeu crucial dans l'élaboration des politiques et dans la conception des systèmes liés aux services publics.

Cette démarche sera basée sur des données concernant l'utilisation des ponts, obtenues auprès des conducteurs automobiles de la grande région de Montréal, via l'enquête téléphonique origine-destination de grande envergure ayant eu lieu en 2003. L'exploitation de cette information pour l'analyse des questions d'équité est rendue possible grâce à l'adoption d'une approche totalement désagrégée.

Les modèles de réseau de transport ont d'abord été développés dans le domaine de la recherche opérationnelle. Un réseau est normalement représenté selon la théorie des graphes. Les segments de rue sont transformés en liens orientés et les intersections deviennent des nœuds. Les liens ont toujours un attribut de coût et souvent un attribut de capacité. Certains nœuds servent d'origine et de destination à la demande, qui est affectée au réseau selon un algorithme spécifié. Un de ces algorithmes fondamentaux est le calcul du chemin le plus court (Bellman, 1958; Dijkstra, 1959).

Dans le cas spécifique de l'affectation des automobiles privées sur un réseau routier, le principe de l'équilibre descriptif de Wardrop (1952) sert d'hypothèse de base. Selon ce principe, chaque conducteur choisit son chemin d'une façon qui minimise son propre temps de parcours. La fonction objective qui représente ce principe a été développée par Beckmann, McGuire & Winsten (1956). Frank & Wolfe (1956) ont pour leur part conçu une première méthode pour optimiser cette fonction, qui est sujette aux contraintes de la conservation des débits et de la non-négativité des volumes. Ces méthodes et ces concepts ont été incorporés dans les logiciels commerciaux qui sont actuellement déployés dans le domaine de la planification du transport urbain et qui forment une partie intégrale du paradigme des quatre étapes.

La méthode de planification en quatre étapes séquentielles (Martin & McGuckin, 1998) est née au milieu du dernier siècle, coïncidant avec l'émergence des premiers ordinateurs, mais avant le début de « l'âge de l'information ». Par conséquent, la méthode repose sur des représentations simplifiées de l'offre et de la demande, qui sont mathématiquement démontrables et qui ne dépendent pas d'une grande quantité de données. Les quatre étapes sont : la génération de la demande, la distribution de la demande, le choix modal et l'affectation au réseau. Il existe pourtant une autre étape préalable aux quatre étapes principales, soit la définition d'un système de zones.

Les conséquences de baser une simulation de réseau sur un découpage territorial arbitraire sont nombreuses. En premier lieu, le découpage apporte un biais d'agrégation à l'analyse. Les résultats de la simulation dépendent de la taille et des frontières de ces zones. Deuxièmement, l'agrégation cause une perte importante d'information en remplaçant une distribution de valeurs par une moyenne. Troisièmement, l'existence d'un système de zones implique la construction de nœuds et de liens artificiels dans le réseau, soit les centroïdes et leurs connecteurs. L'emplacement et la configuration de ces éléments ont une influence non-négligeable sur les résultats d'un modèle d'affectation.

La procédure séquentielle classique et les modèles d'affectation qui en font partie ont souvent été critiqués pour leurs hypothèses simplistes et leur manque de réalisme par rapport à la représentation du comportement humain. En réponse à ces critiques, des modèles de plus en plus complexes au niveau algébrique ont été développés, mais demeurent ancrés dans le paradigme agrégé des quatre étapes.

La troisième étape de la procédure séquentielle classique (le choix de mode) se distingue par sa nature désagrégée. Le type de modèle le plus souvent adopté pour prévoir la part du marché d'un mode de transport en particulier est le logit multinomial de McFadden (1974). L'estimation de ce type de modèle exige des données concernant les choix des individus. Il faut aussi être en mesure d'identifier et de décrire toutes les options disponibles pour chaque voyageur.

Les modèles logit font partie de la classe de modèles d'utilité aléatoire. Des variantes de ces modèles ont été appliquées aux choix d'itinéraire des conducteurs sur un réseau routier

(Ramming, 2002; Frejinger, 2008). Ces méthodes, basées sur la théorie probabiliste du choix discret, représentent une approche assez distincte des algorithmes d'affectation routière normalement appliqués au sein de la procédure à quatre étapes.

L'analyse des erreurs constitue un aspect intéressant des modèles désagrégés de choix. Les prédictions d'un modèle peuvent être comparées aux choix en utilisant une matrice de confusion. Cette matrice permet l'indentification et la caractérisation de certains marchés qui sont mal représentés par le modèle (Spurr & Chapleau, 2007). Une analyse des caractéristiques de ces marchés constitue une méthode systématique pour identifier des lacunes du modèle.

Développée à l'École Polytechnique de Montréal, l'approche totalement désagrégée de la planification des systèmes de transport (Chapleau, 1992) offre peu de ressemblance face à la procédure séquentielle classique. Elle est basée sur une enquête téléphonique d'environ 5% des ménages dans la grande région de Montréal. Cette enquête contient de l'information détaillée sur le comportement des voyageurs pris individuellement. L'unité irréductible de toute analyse est le déplacement et ses attributs. Les attributs de la personne effectuant le déplacement et les attributs du ménage auquel elle appartient sont conservés pendant tout le processus de modélisation.

Parmi les attributs du déplacement, on retrouve l'origine, la destination (géo-référencées au mètre près) et une description partielle du chemin emprunté. Pour les déplacements effectués par transport en commun, cette description inclut les lignes de transport empruntées, alors que pour les déplacements automobiles, elle inclut les autoroutes et les ponts. L'information sur les itinéraires de transport en commun est utilisée pour valider le modèle d'affectation. L'information sur les ponts et les autoroutes n'a jamais été exploitée aux mêmes fins dans un modèle d'affectation routière.

L'approche totalement désagrégée ne dépend d'aucun système de zone. L'agrégation est toujours possible et souvent souhaitable, mais seulement lorsqu'elle est appliquée aux résultats de l'analyse. Cette propriété de l'approche permet une évaluation de nombreux phénomènes, incluant l'équité, sur toutes leurs dimensions.

L'équité en transport fait référence à la manière dont le public est traité par le système de transport. La majorité de la recherche dans ce domaine concerne les populations défavorisées et l'analyse de leurs besoins de mobilité (Forckenbrock & Sheeley, 2004; Litman, 2006). D'autres volets de recherche impliquent l'équité, même si le concept n'est pas toujours explicite. Par exemple, l'implantation d'un système de péage routier pourrait être considérée comme étant équitable, où chaque voyageur paie le « vrai » prix de sa consommation, soit l'internalisation des coûts de congestion. Par contre, un péage peut également être considéré comme un fardeau supplémentaire pour les ménages à faible revenu. L'utilisation des fonds générés par le péage implique elle aussi des questions d'équité.

Le problème de l'équité en transport comporte aussi une dimension géopolitique. Puisqu'il s'agit d'une infrastructure publique, le réseau routier est un puissant mécanisme de redistribution. Une population porte le fardeau des coûts et une autre jouit des bénéfices. L'effet de cette redistribution est équitable dans la mesure où ces deux populations sont les mêmes. Dans plusieurs grandes agglomérations urbaines, un phénomène d'évasion fiscale est observé où chaque municipalité cherche à bénéficier d'une infrastructure sans avoir à en payer les coûts. La victime principale de ce comportement est généralement la ville centrale, qui fournit une proportion importante des services aux banlieues. Cette distorsion fiscale a été quantifiée et rectifiée en partie pour le cas du transport en commun dans la grande région de Montréal, grâce à une analyse de consommation basée sur l'approche totalement désagrégée (Chapleau, 1995).

La ville de Montréal se trouve sur une île liée au réseau routier nord américain par quinze ponts. Ces infrastructures ont une influence importante sur la distribution du trafic sur le réseau régional. Quatre d'entre eux appartiennent au gouvernement fédéral, les onze autres au gouvernement provincial. Dix des quinze ponts ont une classification fonctionnelle d'autoroute alors que les autres sont plutôt des rues artérielles. À des fins de synthèse des résultats, les ponts sont groupés dans quatre lignes écran qui forment ensemble un cordon autour de l'île.

L'enquête origine-destination de 2003 contient 33 000 déclarations de pont. Pour les fins de cette recherche, l'analyse se limitera à l'heure de pointe du matin (6h à 9h) durant laquelle on retrouve environ 9 000 déclarations. Certains déplacements impliquent l'usage de deux ponts. Ceux-ci ont été exclus de l'analyse. De plus, la structure de l'enquête permet d'attribuer les caractéristiques d'un déplacement, d'une personne, d'un ménage et d'un territoire à un pont en particulier, si la déclaration de pont est crédible. La validation de ces déclarations est donc une étape primordiale.

La validation des déclarations de pont est effectuée en estimant deux modèles d'affectation : un modèle de simulation et un modèle de validation. Le modèle de simulation est une affectation tout-ou-rien de chaque déplacement au chemin qui offre le temps de parcours minimal. Le modèle de validation affecte plutôt chaque déplacement au chemin le plus court, mais contraint au pont déclaré dans l'enquête. Une comparaison des temps de parcours et itinéraires générés par les deux modèles permet d'évaluer la plausibilité des réponses. Les déplacements ayant un temps de parcours simulé beaucoup plus petit que le temps de parcours validé sont alors enlevés du sous-échantillon. Parmi les 8 583 déplacements retenus pour l'analyse complète, le modèle de simulation prédit correctement 74.1% des choix de ponts observés. Les prédictions incorrectes peuvent être attribuées à trois phénomènes : l'indifférence, la déviance et l'erreur grossière. Les deux premiers phénomènes sont des éléments incontournables de tout modèle prédictif mais la présence de l'erreur grossière (déclaration erronée, mauvaise codification de la réponse, etc.) doit être minimisée. Malgré le fait que l'identification définitive des erreurs grossières est impossible, le processus de validation suggère qu'elles ne représentent pas plus de 5% de l'échantillon validé.

Pour arriver à construire un modèle de circulation, deux composantes sont nécessaires, soit l'information sur la demande et l'information décrivant l'offre. Un réseau de 104 000 liens et 70 000 nœuds a été construit à partir d'un réseau de simulation développé à l'École Polytechnique de Montréal et d'un réseau numérique fourni gratuitement par le programme GEOBASE du gouvernement du Canada. Le réseau de modélisation comprend deux structures hiérarchiques parallèles : une hiérarchie fonctionnelle et une hiérarchie de juridiction. Toutes les rues de la grande région de Montréal sont incluses dans ce réseau. La vitesse de chaque lien est basée sur sa classe fonctionnelle (autoroute, artère, rue locale) et sur l'observation directe. Puisque l'échantillon de la demande est limité aux déplacements empruntant un pont, ceux-ci sont les seuls liens du réseau auxquels une capacité plausible a été affectée. La capacité de chaque pont a été estimée à partir de la demande observée et du nombre de voies pour chacune des directions.

Afin de conserver la structure désagrégée de l'information détaillée de l'enquête origine-destination, il est nécessaire d'estimer des modèles de choix individuel. Le modèle logit multinomial représente donc une option valable. Le modèle logit est un type de régression linéaire généralisée où la variable dépendante est un choix représenté par les valeurs 0 ou 1, dans ce cas-ci le choix de pont. Plusieurs formulations sont proposées et une matrice de confusion est utilisée afin d'analyser les erreurs et la puissance prédictive de chacune.

Ces trois modèles réussissent à prédire correctement 75% des choix de pont observés. L'utilisation des ponts non-autoroutiers est particulièrement difficile à modéliser, mais les variables incorporant la congestion routière ajoutent peu à la puissance prédictive du modèle. Les caractères différents des réseaux routiers montréalais et des banlieues ont une influence significative sur le choix de pont. En général, un kilomètre parcouru sur l'île de Montréal est associé à une désutilité plus grande qu'un kilomètre parcouru hors-île. Les ponts qui font partie du réseau autoroutier semblent être préférés à ceux qui n'en font pas partie. Sommairement, ces modèles permettent de déduire certaines caractéristiques de l'offre à partir des attributs de la demande. L'estimation des indices de consommation et la construction des itinéraires complets à partir des prédictions des modèles logit sont possibles grâce à une affectation tout-où-rien contrainte au pont prédit par le modèle.

Les modèles de choix désagrégés permettent de conserver les attributs des déplacements et des voyageurs pendant l'affectation de la demande sur le réseau. Par contre, ces modèles ne permettent pas de représenter de façon explicite le caractère dynamique d'un réseau routier urbain. La congestion sur les autoroutes et les ponts est un phénomène qui varie dans le temps et selon le niveau de la demande pour ces infrastructures. Par ailleurs, le réseau urbain est contrôlé par des feux de circulation qui ont une influence importante sur la vitesse moyenne des véhicules, et par conséquent sur le temps de déplacement. Le logiciel « open-source » TRANSIMS est un outil qui incorpore ces phénomènes et qui conserve la structure désagrégée de la demande pendant la procédure d'affectation. De plus, le logiciel enregistre les attributs microscopiques de chaque itinéraire simulé, offrant donc la possibilité d'examiner l'importance des attributs des différents types de rues et de mouvements sur le choix d'infrastructure d'un conducteur.

TRANSIMS est utilisé pour démontrer une approche expérimentale à l'affectation de la circulation routière. Un réseau fictif de 4 000 feux de circulation est développé et un algorithme itératif lui est appliqué pour représenter une congestion dynamique. Au cours des 10 itérations de la simulation, il est possible d'observer les changements de chemin et de pont résultant de la congestion simulée. D'ailleurs, la puissance prédictive du modèle est évaluée à la fin de chaque itération. Les résultats de cet exercice montrent une variation importante des taux de prédiction juste pour ce qui est des ponts desservant une clientèle non-captive. Malgré cette variabilité, le taux global de prédictions justes demeure stable au cours des 10 itérations, soit près de 75%.

Cette stabilité découle du fait qu'une amélioration du taux de prédiction lié à un pont cause normalement une dégradation de ce taux pour un pont voisin.

L'analyse du transport équitable exige une définition quantitative des coûts et bénéfices. À cette fin, les coûts de transport sont associés à la provision de service, tandis que les bénéfices sont associés à la consommation. Dans la grande région de Montréal, les services de transport routier sont fournis par les gouvernements municipal, provincial et fédéral et sont consommés par les ménages de cette région.

Les coûts associés à l'offre de services de transport se présentent de différentes façons. Les coûts directs monétaires de la construction, de l'entretien et de l'exploitation sont assumés par le gouvernement auquel appartient l'infrastructure en question. Les autres coûts moins facilement quantifiables, tels la pollution et le bruit, sont assumés par les personnes qui vivent, travaillent ou évoluent près de cette infrastructure. Tous ces coûts sont considérés comme étant proportionnels au nombre de véhicules-kilomètres associés à celle-ci.

Il existe également des coûts associés à l'utilisation du réseau routier qui sont assumés par le conducteur lui-même. Les coûts monétaires découlent de l'achat, de l'entretien et de l'exploitation d'un véhicule. Le principal coût non-monétaire est le temps consacré au déplacement. Dans une perspective d'équité, ce coût n'est pas intéressant en soi. Par contre, le coût par kilomètre parcouru varie de manière importante en fonction du déplacement. Ce coût marginal de déplacement est calculé comme le rapport du nombre de véhicules-heures sur le nombre de véhicules-kilomètres. Cette quantité est l'inverse de la vitesse moyenne et il est représentatif du niveau de service reçu par le conducteur et offert par l'exploitant de l'infrastructure.

Les résultats du modèle de validation permettent l'estimation des effets redistributifs associés à l'usage des ponts de Montréal. Le calcul des coûts et des bénéfices est donc basé sur un comportement observé et il est rendu possible par l'approche totalement désagrégée, qui permet l'association d'une infrastructure particulière à des usagers, ménages et territoires spécifiques. Les unités d'analyse sont les 100 secteurs municipaux qui composent la grande région de Montréal. Ces secteurs peuvent être agrégés en 5 sous-régions : l'île de Montréal, la Rive sud, Laval, la Couronne sud et la Couronne nord. Le nombre de véhicules-kilomètre consommés et offerts est calculé pour chaque secteur et sous-région. Une comparaison de ces deux quantités permet de déterminer dans quelle mesure une région paie un coût qui soit proportionnel aux bénéfices qu'elle reçoit. Les résultats de l'analyse montrent que les secteurs de l'île de Montréal ont des ratios bénéfice-coût nettement inférieurs à 1, alors que les secteurs des deux couronnes ont des ratios supérieurs à 1. Les secteurs de Laval et la Rive sud subissent quand à elles des coûts proportionnels aux bénéfices reçus.

Le deuxième volet de l'analyse examine le temps par kilomètre parcouru offert par les gouvernements municipaux et provincial. L'application du modèle de validation révèle que les infrastructures du gouvernement provincial ont un coût marginal de déplacement qui est moins important que le coût marginal exigé par l'usage de l'infrastructure municipale. Autrement dit, les vitesses moyennes sur les infrastructures provinciales sont plus élevées que les vitesses moyennes sur les infrastructures municipales. Cet approvisionnement de service à prix réduit encourage des déplacements de longue durée sur des infrastructures localisées très loin du domicile. Donc, les infrastructures provinciales contribuent d'une manière importante aux distorsions évidentes dans la distribution des coûts et bénéfices parmi les municipalités de la grande région.

L'exercice décrit dans la section précédente a démontré que les infrastructures à caractère national (les autoroutes et ponts provinciaux, et les ponts fédéraux) contribuent à une distorsion importante dans la redistribution des coûts et bénéfices du transport routier. La concentration des coûts dans la ville centre et la concentration des bénéfices dans les régions périphériques stimulent un comportement d'évasion fiscale, qui s'exprime par la croissance fulgurante des populations banlieusardes et par la stagnation permanente de la population de l'île. L'approvisionnement des infrastructures routières lourdes par les gouvernements centraux contribuent donc à l'étalement urbain.

Une solution équitable, déjà implantée dans les systèmes de transport en commun de la grande région de Montréal, serait d'exiger que chaque municipalité paie non seulement pour les infrastructures dont elle est responsable, mais aussi pour la consommation de ses résidents sur les infrastructures des autres municipalités et des autres gouvernements. Un tel mécanisme impliquerait un péage urbain qui serait plus acceptable que l'internalisation des « coûts de la congestion » par l'implantation de frais d'utilisation sur les infrastructures majeures.

L'acceptabilité politique de la solution repose sur la crédibilité de la méthode de calcul et sur la qualité de l'information sur laquelle elle est basée. Une dernière composante de la recherche consiste en l'estimation des indicateurs de redistribution à l'aide les résultats du modèle de simulation. La différence entre ces résultats et ceux du modèle de validation sont mineures. L'approche totalement désagrégée appliquée au système routier aurait donc une certaine valeur comme outil de prévision dans les études de transport équitable.

Cette recherche a démontré la faisabilité de mesurer la distribution des coûts et bénéfices du transport routier par la construction d'un modèle de circulation régional totalement désagrégé. Les déclarations des ponts contenues dans une enquête origine-destination ont été validées, puis plusieurs modèles de simulation qui conservent les attributs des voyageurs ont été estimés. Tous les modèles, y inclut ceux qui représentent les phénomènes de congestion, réussissent à prédire environ 75% des choix de pont observés. Un modèle de validation qui simule le chemin incorporant le pont déclaré a également été développé, ce qui a permis la comparaison avec les itinéraires des modèles de simulation. Cette comparaison désagrégée a facilité la compréhension des différents facteurs qui influencent le choix de pont.

Des définitions quantitatives de coûts et bénéfices du transport routiers ont été élaborées. Celles-ci ont été analysées l'aide du modèle de validation et ont démontré que les coûts et bénéfices associées à l'utilisation des ponts ne sont pas distribués équitablement entre les différentes municipalités de la grande région de Montréal. Les banlieues accumulent les bénéfices tandis que la ville centre porte le fardeau démesuré des coûts. Cette distorsion est attribuable à l'exploitation d'infrastructures lourdes par les gouvernements nationaux (provincial et fédéral). L'existence de celles-ci facilite un comportement d'évasion fiscale qui contribue à l'étalement urbain. Le redressement de cette situation dépendrait d'une méthodologie crédible et impartiale permettant de calculer la répartition équitable des coûts associés à l'utilisation des grandes infrastructures par les diverses municipalités. La méthode développée dans cette recherche, étant basée sur un comportement réel et observé, pourrait être appliquée à cette fin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The Jacques-Cartier Bridge is a magnificent construction of stone and steel almost two kilometres long stretching over the St. Lawrence River and connecting downtown Montreal with the adjacent city of Longueuil. From the perspective of a civil engineer, the bridge, with its arced profile and elaborate truss structure suspended more than 30 metres above the water, is a life-giving artery to a large city and a monument to human achievement. Such structures are what motivate engineers to pursue careers in the field, secure in the knowledge that they will make a tangible and lasting contribution to civilization. Society at large is also inspired by these grandiose construction projects which represent progress, modernity, prosperity and power. Transportation projects are especially seductive because of the opportunities they provide to apply state-of-the-art technology toward the liberation of people through high-speed travel. Each improvement in transport service is a metaphorical bridge – a new shortcut which reduces the cost of travel between multiple locations. Nevertheless, progress, however defined, always comes with a cost and this cost is not necessarily smaller than the presumed benefits. Intense debate inevitably arises whenever a large infrastructure project is proposed because questions of "costs" and "benefits" are essentially questions of semantics. The words mean very different things to different people. It is this fog of subjectivity which creates a role for the transportation engineer, who is called upon to construct an objective framework.

The objective framework for the analysis of transportation systems in large cities is known as "urban transportation planning". It provides a basis for collaboration between politicians, engineers, architects, professional planners, and others. Since it involves a wide array of disciplines and a very large number of individual participants, urban transportation planning has developed its own culture which, like all cultures, is in a constant state of flux. Its original preoccupation was the optimization of the complex system formed by urban transport infrastructure and the vehicles that use it. New preoccupations include road pricing, sustainable development, environmental justice and social equity. These concepts are not easily incorporated into a quantifiable objective function and therefore present a challenge to an engineer. It is an appropriate challenge, since, in the words of Manheim, the methodological challenge faced by the transportation analyst is to conduct "a systematic analysis in a particular situation which is valid, practical, and relevant, and which assists in clarifying the issues to be debated" (Manheim, 1979). This thesis represents an effort toward the clarification of the new planning paradigm for the purposes of quantitative analysis of road networks.

Integral to the new planning paradigm is an increased emphasis on the role of non-automobile forms of transport. A broad but diffuse consensus exists among the political class and the population at large that the automobile imposes unreasonably large costs on society. These costs take many forms and include: injuries and deaths caused by collisions, paralyzing traffic congestion in urban areas, noise, severe air pollution, climate-altering greenhouse gases, and increasingly violent conflicts over disputed petroleum supplies. Finding innovative ways of reducing car use has become a motivating force behind research into new analysis methods as a result. The persistence of this trend can be attributed to the ease with which the costs of car travel can be quantified and their distribution described. The police and insurance companies keep detailed accident records, air quality is measured directly, a price has already been put on airborne carbon, and the costs of armed conflict are well-documented. Calculating the benefits of car use, on the other hand, is more complicated since travel is motivated by a desire to participate in non-travel activities. In addition, determining the distribution of benefits is difficult because there are few standard methods for precisely identifying the beneficiaries of a particular facility.

Urban transportation planners specializing in road systems have at their disposal many sophisticated tools for the scientific analysis of transportation problems. Of particular interest to the present research is the class of simulations known as traffic assignment models. Traffic assignment methods have evolved considerably since their inception in the middle of the last century and now incorporate all the detailed elements of road traffic. A thorough review of the literature demonstrates, however, that these tools are designed primarily to produce performance indicators for network elements. They are not particularly well-suited, however, to the identification of system *users*. One reason is that the tools widely used in practice today were developed at a time when detailed information on driver behaviour was exceedingly difficult to obtain. While this is still the case in most places, it seems unlikely to remain so. The quantity of information being collected through modern communication technology was unimaginable until very recently. From a technological perspective, it has become a trivial matter to track an individual person or vehicle on a transportation network. This research aims to develop a methodology for exploiting such information (in the event that it becomes available to urban transportation planners) with the purpose of identifying the clientele of road infrastructure facilities.

Once users have been identified, it is possible to gain some insight into the distribution of the benefits and costs of associated with a road network among various population groups. More specifically, a dynamic, detailed and totally disaggregate treatment of infrastructure use permits an equity analysis of road transportation. Here, the term “equity” refers to fairness and can be applied to a transportation system to the extent that those who benefit from it (or parts of it) are not necessarily the people who pay its costs. The ability to distinguish between the two groups seems essential in a discussion of topics such as road-user charging or increased investment in improved transit service.

The Jacques-Cartier Bridge is one of fifteen bridges which provide road access to the island city of Montreal. These high-capacity transport facilities play a central role in the daily operation and the long-term development of the Greater Montreal Area. They are also among the most congested road segments in the region. This thesis employs a large sample of revealed preference information on the choice of major bridge in the Greater Montreal Area. The data are a subsample of a detailed travel survey which interviews roughly 5% of the region’s households. In addition to their choice of bridge, respondents are also asked to provide personal and household attributes, as well as supplementary information on the trip itself such as the origin, the destination, the purpose and the time of departure. These data are initially employed to examine in detail the factors that cause drivers to choose a particular facility among several possible options. The application of the totally disaggregate analysis approach means that multiple aspects of driver behaviour are analysed independently. Totally disaggregate network simulation models are subsequently developed to estimate measures of consumption, supply, costs and benefits induced by the existence of the major bridges. The network used for the simulation is extremely detailed, containing in excess of 100,000 links and the adopted simulation methods are performed using sophisticated modelling platforms recognized as being at the cutting edge of transportation modelling technology. Revealed preference data describing usage patterns represents not only a unique opportunity to study the costs and benefits conferred by large road infrastructure, but also a promising environment for the development of new regional traffic models that permit the just evaluation of such questions.

This document is divided into 4 chapters. The first chapter is a review of current methods in traffic assignment modelling. The second chapter describes an information-based totally disaggregate approach to the traffic assignment problem. The third chapter discusses the application of this approach to the question of equity in road transport. The last section is a summary conclusion.

CHAPTER 1 REVIEW OF CURRENT PRACTICE

This research was conducted within a sub-branch of civil engineering known as transportation engineering. Transportation engineering can be further subdivided in numerous ways, but one important distinction should be made between the design of infrastructure *elements* and the design of infrastructure *systems*. The former is concerned with geometry and vehicle kinematics and is applied to the conception of physical components of the road system such as rights-of-way, intersections, signage, signals and lighting. The design of infrastructure systems concerns the interaction of vehicles with each other and with the built environment. While the design of vehicles and roadways is mostly deterministic, the behaviour of large numbers of travellers acting independently from each other is often difficult to predict. The branch of transportation engineering concerned with analyzing this complex human system is often known by the more general term “urban transportation planning”.

Urban transportation planning has existed as an independent discipline for at least 50 years and was developed with the aim of applying some scientific rigour to the process of deploying urban transport infrastructure. While the science has progressed continuously over time, it rarely carries much weight in the decision-making process. At best, the opinion of the transportation planner is just one of many elements to be considered. Indeed, the transportation planner is so dispensable that many smaller jurisdictions do not bother employing one. The reasons for this state of affairs are not mysterious. In 2007, the Transportation Research Board released a document entitled “Metropolitan Travel Forecasting: Current Practice and Future Direction” (Transportation Research Board, 2007). Based on a comprehensive review of the state of the art in the United States, the report presents several findings. In particular:

“Shortcomings of conventional forecasts are also related to poor technical practice in the use of models.”

The report provides several examples of “poor technical practice” including inadequate data and a lack of quality control. Moreover, there are countless examples of forecasts being manipulated to serve particular interests despite the objections of trained professionals – a sure indication of a fallible methodology. Yet the foundations of urban transportation planning are not unsound. At its heart is the notion of modelling – constructing a virtual representation of the system for the purposes of predicting how the system will respond to specific interventions. Every science is based upon a set of models which have been empirically demonstrated to be valid in a specific context. A proper validation process is difficult to achieve, however, when the system being modelled is composed of human beings interacting with each other. Therefore, before discussing in detail the methods developed in urban transport planning, the review of current practice begins with a discussion of models of man, the collection of data, and construction of knowledge.

1.1 "Models of man"

A model can be thought of an isomorphism – a formal system whose theorems have meaning in the real world even though the system itself is artificial and independent of reality. The extent to which humans rely on isomorphism in their cognitive processes cannot be overemphasized. Each person's understanding of reality is not limited to their perception at a given moment in time. Experiences are recorded (remembered) and the cumulative effect of these experiences leads us to construct models to help us foresee the consequences of a particular action. Although this hardwired thought process leads us to frequently make mistakes (incorrect predictions), without it, life would be extremely difficult since each situation would be entirely new and unique. The act of learning would be impossible.

One of the shortcomings of transportation planning as a science is that functional isomorphisms are rare. Mathematical models abound. But their relationship to reality remains uncertain. Most urban transportation planning models are derived from the conventional wisdom of economic theory which holds that human beings have advanced cognitive abilities which allow them to optimize their own particular situation using a process of complex reasoning. An alternative view was advocated by Herbert Simon who instead proposed that the complexity lies not in the human decision-making process, but rather in the environment in which these decisions are made (Simon, 1996). Simon also suggested that humans never have complete information and are not generally able to perform sophisticated calculations in a short time period. Rather than choosing the optimal course of action, we choose one that is "good enough", given the information available at the time. This theory is encapsulated in the principle of "bounded rationality" (Simon, 1957). For purposes of transportation planning, it is therefore incumbent upon the travel behaviour analyst to obtain, to the greatest possible extent, a complete picture of the context which informs traveller choices. The planner should also take care to avoid attributing an unrealistic burden of computation and profound reasoning to his human subjects.

The primary motivation behind the present study is to identify the clientele of major road infrastructure. A key step toward realizing this goal is the construction of a reliable model of the route choice process of drivers. If Simon is to be believed, the decision process itself should be fairly simple. A driver will use a path that he finds satisfactory and he will choose this path from within the limited set of alternatives of which he is aware. The challenge, therefore, is to construct a network model which accurately represents the road characteristics which are relevant to the choice of route.

1.2 The collection of information and the construction of knowledge

The early twenty-first century is justly referred to as the "information age". Advanced systems of perception and digital data storage have led to the construction of formidable repositories of pure information. This development represents a great opportunity for the advancement of knowledge in many domains, but it seems especially important for what Simon called the "artificial" sciences – those that deal with human systems.

Although this opportunity has been readily acknowledged, it has been embraced only cautiously by many disciplines, including transportation planning. One reason is that the availability of information, in addition to presenting a great opportunity, also poses significant challenges. Most notably, the treatment of millions or billions of observations requires a set of skills held by only a fairly small number of specialists. The vast majority of civil engineers, for example, receive no training in how to process large quantities of data automatically and efficiently. A second reason is that digital information is by definition discrete. This means that analysts are required to look beyond traditional algebra which deals only with continuous quantities. Decreasing our reliance upon such a compact, efficient and elegant system of thought with something more complicated will not be an easy task.

Another challenge is related to the fact that raw data represent a chaotic analytical environment. While a machine and its associated operating system are very good at measurement and calculation, they necessarily have a limited power of inference. As a result, there are many discrepancies between what is gathered as information and what that information is supposed to represent. A good example in transportation planning is smart card data. The card reading machines ostensibly measure the number of passengers entering the system at particular locations, when in fact they measure only the number of cards with which they came into contact. The difference is subtle but important. Information alone is by no means equivalent to knowledge.

The construction of knowledge implies the treatment of raw data using a method that gives meaning to the collected information. This complex task is essential for the subsequent development of usable and useful models. Generally speaking, there exists a large array of different approaches but two concepts in particular are emphasized here. The first is the object-oriented framework in which data are classified as being attributes of well-defined entities. If the relationships between entities can be properly represented then a significant part of the knowledge construction process is already accomplished. The second concept is visualisation, which is often a far more effective means of communicating ideas than raw text or numbers. The process of designing appropriate visualisation tools also contributes greatly to the meaningful structuring of information. An example of the simultaneous application of these two concepts is found widely available Geographic Information Systems (GIS) which have come to play a central role in transportation planning and analysis. Basic applications are described in a paper by Miller (H. J. Miller, 1999) but recent examples of GIS usage in transportation abound ((Buliung & Kanaroglou, 2006; Stopher, 2004b; Trépanier, Chapleau, & Allard, 2002), for example).

In effect, the construction of knowledge is the construction of a conceptual model. A validated model is used to explain the relationships between numerous measurable quantities. Moreover, the values of some of these quantities can be calculated based on the measurement of others if the model is sufficiently isomorphic with reality. Traffic assignment models, for example, describe the relationship between measurable quantities such as road capacity, the volume of traffic on links, the average speed of traffic and the demand for auto travel between origin-destination pairs. It is an economic model which seeks equilibrium between supply and demand. Once the parameters of a traffic assignment model have been estimated, the supply and demand are used as known inputs and the equilibrium condition is the desired output. All three must be simultaneously observed to determine the model parameters. The next three sections describe, respectively, demand data, supply data and data relating to the point of equilibrium which we call transaction data.

1.2.1 Data describing road transport supply

A regional model of urban road traffic is based on a representation of road infrastructure. Two classes of data are required: connectivity data and level-of-service data. Connectivity data is necessary for determining which paths are possible within the network. Level-of-service data is used to construct path attributes for the purposes of comparison.

Complete network connectivity data have been collected by numerous public and private organizations which have produced very detailed spatially referenced databases describing the geographic positioning of road networks elements. These digitized networks, which are well-suited to analysis using standard GIS, are now widely available for relatively little cost, especially when compared to the costs of collecting such information manually.

Level-of-service information relates to the speed and capacity of a network component. While information on the connectivity of streets can be compiled using, for example, aerial photo technology, road segment speeds and capacities depend on microscopic attributes of the road in question, particularly signage and signalling systems. An additional complication arises from the fact that the level-of-service offered by a road facility also depends on the volume of traffic using it (see section 1.3.1.3) and this simultaneous estimation of network performance and travel demand lies at the heart of the traffic assignment problem.

While early traffic assignment models required only a simplistic and approximate description of the regional network, current methods require detailed information about microscopic elements of the road system such as traffic lights, stop signs, turn restrictions, turning bays, merges and diverges. The collection of all this information for an entire metropolitan area represents a significant challenge, but the task has been rendered almost manageable by contemporary data collection, storage and communication technology. In principle, the authorities responsible for the operation of these systems possess inventories of their own transportation hardware but these repositories are frequently not in digital format and so manual codification is still necessary.

Advancements in information technology continue to offer ever more powerful means for coding networks at the microscopic level. A well-known example of this type of tool is the Streetview application developed by Google. Made available free of charge to anyone with an internet connection, Streetview provides panoramic photographs of thousands of kilometres of urban and rural streetscapes around the globe. Although structured automated queries cannot be made on this database, it nevertheless represents a rich repository of information describing the location of traffic signals, signage, parking regulations, lane configuration, intersection geometry and road surface conditions.

1.2.2 Data describing road-transport demand

Historically, demand for transport service has been evaluated using direct observation of vehicle or passenger volumes at particular locations. Most traffic assignment models are calibrated using vehicle count data. While they are frequently the only available source of demand data, they contain no behavioural information and their reliability is often questionable. The quantity of vehicles that passes a given location during a given time interval reveals nothing about the origins and destinations of the vehicles or their passengers. Moreover, the counts themselves, when not undertaken by humans, are performed by machines which must be properly calibrated, maintained and monitored by humans. In addition, the high cost associated with installing equipment or sending people on to the street for several hours means that counts at a particular location are performed on one day only. As a result, any statistics computed using the resulting information cannot be considered statistically significant. Many jurisdictions have installed permanent devices which measure vehicle volumes and speeds, particularly on freeways, but this practice is not common the Greater Montreal Area.

In order to acquire a more precise picture of travel demand, it is necessary to collect information on individual travellers and their travel habits. A long-standing method for investigating people's travel behaviour is to simply ask them using a survey. An important body of research exists which documents the evolution of travel survey methods (Stopher & Stecher, 2006). There are many issues surrounding the design of a travel survey including sample size, sample representativeness, the design of interview questions, and the interview technology. Travel surveys can be classified in numerous ways. Important distinctions exist between panel, cross-sectional and time-series data. Another important distinction is stated preference vs. revealed preference.

A stated preference survey presents respondents with a hypothetical question such as: if there were a Maglev train which could take you from New York to Boston in under an hour, would you use it? The answer, of course, is also hypothetical so a great deal of care needs to be taken in the design of the question and the interpretation of the results (Hess, Rose, & Polak, 2008). There is some debate over the utility of stated preference surveys for behavioural studies.

A revealed preference survey asks people what action they have already taken. An example would be: What mode of transport did you use to go to work yesterday?" Although the veracity of the response cannot be guaranteed, it will at least be based on a presumed statement of fact. Geographic Positioning Systems (GPS) technology provides another form of revealed preference information and is sometimes used to augment travel surveys (Bohte & Kees, 2008; Stopher, FitzGerald, & Xu, 2007).

1.2.3 Transaction data

In classical economic theory, the demand and supply functions for a particular good intersect at the transaction price. It is the point where the consumers' willingness to pay coincides with the cost of production. Under stable conditions, this point represents equilibrium between supply and demand. If the transaction price and the quantity of transactions can be directly observed then the point of equilibrium can be defined.

In travel behaviour models, the concept of equilibrium is more closely linked with a quantity of time than with a quantity of money. Therefore, the out-of-pocket monetary cost of a particular trip is less interesting than the amount of time the trip consumed or, in other words, the amount of time spent on the network for a particular trip purpose. Obtaining simultaneous measurements of volumes and travel times is not simple, but several approximate methods do exist. A few of them are described below.

1.2.3.1 Smart cards

The measurement of a public transit system is generally easier than the measurement of a road system since access to the transit system is much more controlled. Fare collection procedures have often provided a side-benefit of measuring usage patterns in public transit systems. Smart cards offer an especially rich opportunity in this regard. The data itself are very simple – consisting of a time, a card number and an equipment identifier. When a single transaction is combined with other transactions and other sources of data such as GPS tracks of transit vehicles and descriptions of the planned service, it is possible to observe not only the monetary price paid for a trip, but also to impute for a given cardholder the number of kilometres travelled, the amount of time spent travelling, line load profiles and many other quantities as well (Chapleau & Chu, 2007; Chu & Chapleau, 2008).

1.2.3.2 Public transit vehicle probes

Unlike public transit systems, public road networks are usually not designed to collect revenue directly and so information on the road clientele is frequently non-existent. Fortunately, most surface vehicles operated by public transit agencies share the right-of-way with private automobiles and, as a result, they can act as probes to indirectly measure the performance of road systems. In the Montreal context, the planned service of a bus line has been used to visualize variations in travel time over the course of a day on a single urban arterial road (Chapleau & Piché, 2009). To be sure, a public bus interacts with road traffic quite differently from a private car, but attempts have been made to control for phenomena specific to transit operations. Tétrault and el-Geneidy (2009), for example, estimated a model of bus run times as a function of weather conditions, time of day, vehicle type and the number of boardings and alightings.

1.2.3.3 GPS and crowdsourcing

In many jurisdictions, including Montreal, there is a growing interest in finding mechanisms which would facilitate the marginal cost pricing of roads. Geographic Positioning Systems show great potential since, in theory, they would permit a central authority to know where, when and over what distance a vehicle used a precise facility (see for example (Gilliéron & Waegli, 2005; Nazer & Pickford, 2007)). The implementation of such a system would have the added benefit of generating a very detailed database on the routes used by drivers.

Wireless communication devices in vehicles offer the potential to reconstruct road traffic conditions in real time (Barth, 2009) and to simulate average conditions using historical data (Figure 1.1). Comparable methods are used by corporations which provide real-time traffic information to GPS receivers installed in private vehicles (Downs, Chapman, Cahn, & Hersch, 2006). The process of structuring this data to provide coherent information is known as “crowdsourcing”. It provides an approximate image (or series of images) which reveal the evolution of travel times on a large number of network links over time. The resolution of the technology is such that information can be provided over intervals of just a few minutes.

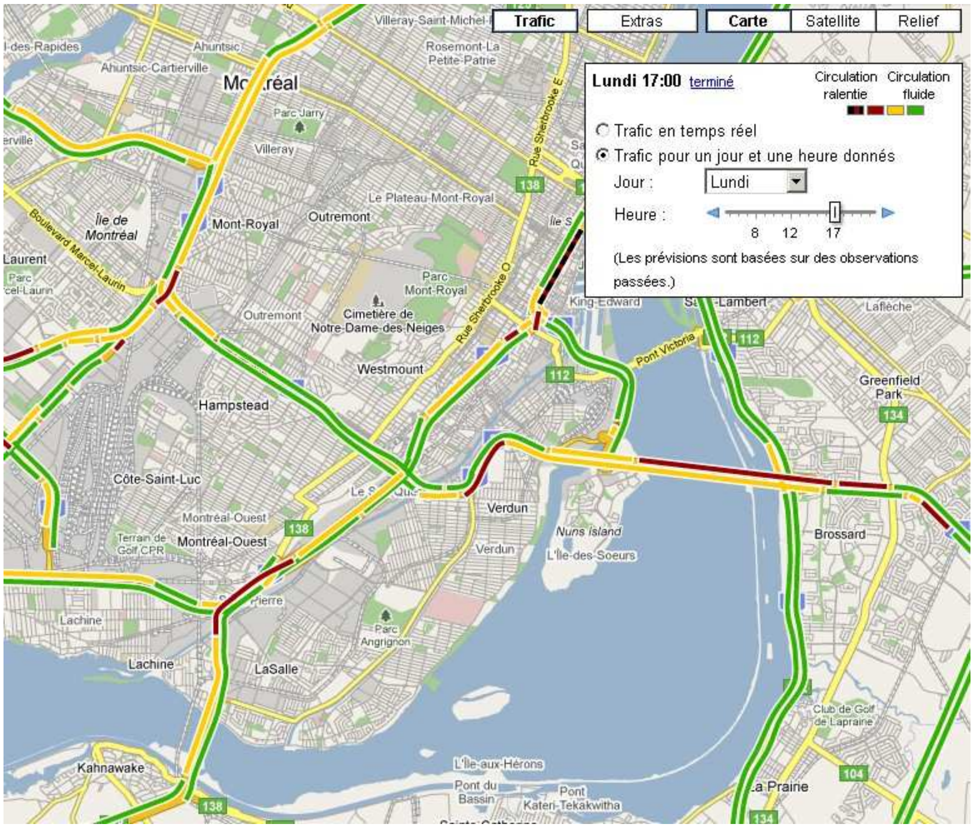


Figure 1.1: Google traffic (maps.google.ca)

1.3 Transportation planning

From an idealized engineering perspective, transportation planning is an exercise in optimization. A particular objective is sought subject to some known constraints. Strategic planning of this sort is common within private transportation firms who have every incentive to minimize costs and maximize revenue. The management of vehicles and personnel through space and over time is essential. This paradigm is also applied to the operation of publicly-owned transit systems because, even though the profit motive is usually absent, the centralized control structure allows for the optimal allocation of limited resources. Mass transit is usually considered a public service however, and as such it is often planned to achieve objectives which are more political than economic. The same can also be said of public road systems, albeit with the added complication of non-centralized control. Most often, the organization responsible for the operation of a road network has no power to route vehicles or to schedule trips. Despite continual advances in the domain of intelligent transportation systems (ITS), a true optimization of the system is impossible.

The planning of public roads, therefore, is less about working toward a clearly defined goal (the optimization of an objective function) than it is about evaluating proposed development strategies. An effective tool for accomplishing this task is the cost-benefit analysis. Unfortunately, many of the impacts of transportation projects – such as air pollution or decreased travel times – are difficult to measure in monetary terms. Moreover, a political environment which evaluates policy based on intent rather than effectiveness serves to reduce the importance of objective measures of performance in favour of symbolic gestures responding to a "categorical imperative" – for example, the reduction of greenhouse gas emissions. The effect of the shift in focus from comparisons of costs and benefits toward the achievement of categorical imperatives has led to the deterioration of the logical decision framework which informs transportation planning. Road network planning in particular has become problematic, since the primary goal of many contemporary transport strategies is to *reduce* the usage of the private automobile (Orfeuil, 2008).

The lack of objective clarity has not prevented the development of a scientific culture of urban transportation planning, as embodied by the classic four-stage model (Martin & McGuckin, 1998). Numerous concepts have been defined and relationships between them have been established. An arsenal of algebraic models of widely varying complexity have been developed and applied. In fact, the increasing sophistication of these algebraic approaches to traffic modelling has been the focus of researchers and practitioners, often at the expense of the empirical approach. Nevertheless, a set of scientific "first principles" describing the observed behaviour of independent vehicles on a roadway does exist. This empirical basis informs the codification of the engineering standards which regulate road design. The components of the four-stage model which deal with automobile traffic incorporate these principles to varying degrees. The next section of this chapter briefly discusses the basic models of traffic. The following section discusses the four stage paradigm and its representation of urban road traffic.

1.3.1 Models of traffic

The movement of multiple vehicles on a road network gives rise to the phenomenon known as traffic. While traffic appears chaotic at the level of an entire network, at the level of individual vehicles the basic kinematic principles are well-documented. An authoritative reference on traffic models is the book by May (May, 1990). Vehicles are typically modelled as particles moving through space at variable velocity and acceleration. The velocity of a given vehicle is dictated by the vehicles in front, the control system, and the surrounding environment. Beginning from a theory of the motion of an individual vehicle, it is possible to develop models of traffic flow which are somewhat analogous to models of fluid mechanics. The short discussion of such models in the next section is based on course notes from Prof. Karsten Baass of École Polytechnique de Montréal (2003).

1.3.1.1 Fundamental Variables

An automobile is characterized by its length, its width, its acceleration (the speed with which it can change its velocity), and the perception-reaction time of its driver. When the analysis is constrained to a single lane, a basic behavioural assumption is that drivers keep a minimal distance between their own vehicle and the vehicle immediately in front. The model can be derived from a formula for the spacing (S) between two vehicles as a quadratic function of vehicle speed:

$$S = L + t_{pr}V + \frac{V^2}{f_2 + g} - \frac{V^2}{f_1 + g} \qquad (1.1)$$

where L is the length of the leading vehicle, t_{pr} is the perception-reaction time of the driver, V is the speed of both vehicles, f_2 is the braking rate of the following vehicle, f_1 is the braking rate of the leading vehicle and g is the longitudinal grade of the roadway. The spacing between two vehicles is determined based on the second driver's own assumptions about f_1 and f_2. An aggressive driver acts on the belief that the braking rate of the vehicle in front (f_2) will be manageably small, thus providing him with sufficient time to brake before colliding with the leading vehicle.

The inverse of vehicle spacing is defined as traffic density (K) which is calculated as an average over a specified length of road (l).

$$K = \frac{N}{l} = \frac{1}{S} \tag{1.2}$$

where *N* is the number of vehicles on the road segment at the moment of observation. The average speed of the traffic stream, *V*, depends on the time, *t*, required by each vehicle to travel over the segment of length *l*. This average, called the space-mean speed, is equivalent to the harmonic mean of individual vehicle speeds (v_i):

$$V = \frac{Nl}{\sum_{i=1}^{N} t_i} = \frac{N}{\sum_{i=1}^{N} \frac{1}{v_i}} \tag{1.3}$$

The average traffic flow rate, *Q*, is computed as the product of *K* and *V*.

$$Q = KV \tag{1.4}$$

This equation relating speed, flow and density forms the basis of the uninterrupted model of traffic flow. It represents the behaviour of vehicles in motion and breaks down when vehicle speed becomes zero. It is easily shown that *Q* is a convex function of *V* and the maximum value of *Q* is defined as the capacity (Q_{max}) of the road segment. Very briefly, equations (1.1), (1.2) and (1.4) can be combined to obtain

$$Q = \frac{V}{a + bV + cV^2} \tag{1.5}$$

where *a*, *b* and *c* are the factored constant terms in the expression for spacing (equation 1.1). The first and second derivatives of *Q* with respect to *V* are:

$$\frac{dQ}{dV} = \frac{\frac{a}{V^2} - c}{\left(\frac{a}{V} + b + cV\right)^2} \tag{1.6}$$

$$\frac{d^2Q}{dV^2} = -\frac{\frac{2a}{V^3}}{\left(\frac{a}{V} + b + cV\right)^2} - \frac{2\left(\frac{a}{V^2} - c\right)^2}{\left(\frac{a}{V} + b + cV\right)^3} \tag{1.7}$$

Since the second derivative of flow with respect to speed is strictly positive, the space-mean speed corresponding to the maximum flow rate is found by setting the first derivative equal to zero. V^* denotes the space mean speed corresponding to the maximum flow rate.

$$\frac{dQ}{dV} \equiv 0$$

$$V^* = \sqrt{\frac{a}{c}} \qquad (1.8)$$

Finally, this optimal velocity is used to find the capacity of the road segment:

$$Q_{max} = \frac{1}{2\sqrt{ac} + b} \qquad (1.9)$$

Numerous functional forms for the speed-density relationship have been proposed but all represent Q as a convex function of V. In classical traffic theory, the function has two distinct regions each of which represents a different traffic regime. The part of the function over which the flow rate increases with traffic speed is the "forced flow" or "congested" regime. The range of speeds over which the flow rate decreases with increasing speed is the "free-flow" regime (Figure 1.2).

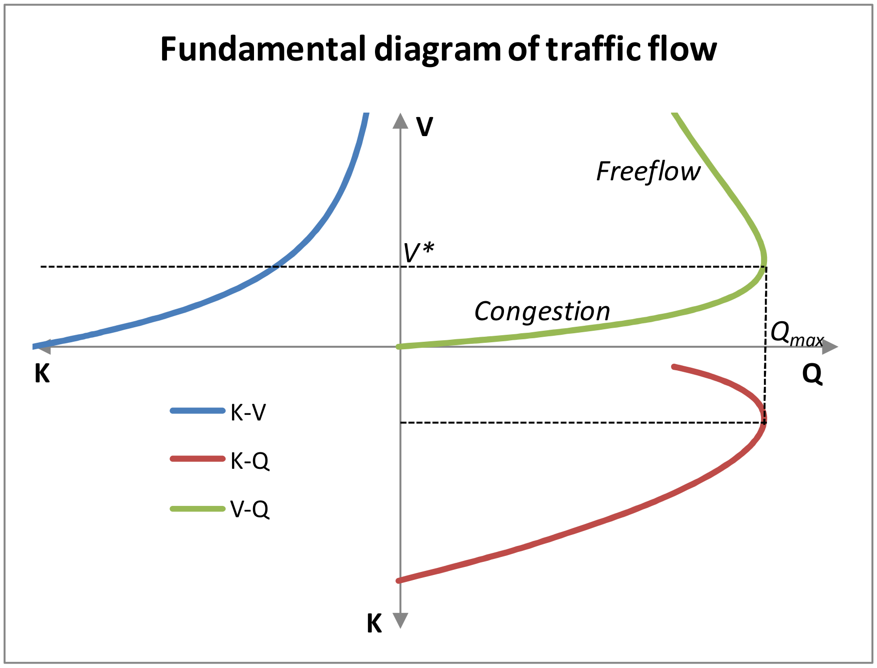


Figure 1.2: Fundamental diagram of traffic flow on an uninterrupted flow facility

The model provides a coherent description of traffic flow on an uninterrupted facility such as a freeway by incorporating vehicle speed, acceleration and braking, all of which are dynamic characteristics. The model does not, however, account for the dynamic variations of network characteristics. Traffic signal phases, accidents on the road, adverse weather and other events distributed over time cause abrupt changes in the capacity of specific network links. These changes interrupt the traffic stream and cause the fluid-flow model of traffic to break down. An alternative analysis is approach is necessary.

1.3.1.2 The Queuing Model

Queuing models of roads employ variations on the first-in-first-out (FIFO) queuing model. In a queuing model, the capacity of a facility is an exogenous parameter defined as the sustained maximum rate at which a facility can process vehicles. The demand for the facility is also represented as a rate. If the demand exceeds the capacity, then a queue will form. The queue will continue to lengthen until demand falls below capacity. The demand and arrival rates can vary with time. The queue length and associated waiting times will vary with time as well. The associated algebra allows for the estimation of important performance indicators such as average queue length, average time spent in the queue, and total delay incurred at the facility. Different algebraic formulations exist to handle deterministic and probabilistic assumptions about vehicle arrivals and facility capacity.

The queuing model is easily constructed using a queuing diagram (Figure 1.3). The diagram is a two-dimensional graphic with time on the horizontal axis and the cumulative number of vehicles on the vertical axis. Two functions are plotted on the graph: the arrival function and the departure function. Both functions by definition are non-decreasing. The length of the queue in terms of number of vehicles at time t, $N(t)$, is found through the expression

$$N(t) = \max\{0, A(t) - D(t)\} \qquad (1.10)$$

where $A(t)$ and $D(t)$ are the cumulative number of arrivals and departures, respectively, at time t. The arrival rate is specified according to the temporal distribution of incoming traffic. There are numerous functional forms that are used in practice. The departure function depends upon the existence or non-existence of a queue at time t.

$$D(t) = \int_0^t d(u)du \qquad (1.11)$$

Where

$$d(t) = \begin{cases} s \text{ if } N(t^-) > 0 \\ a(t^-) \text{ otherwise} \end{cases}$$

and s is the maximum service rate of the facility, $a(t^-)$ is the arrival rate just prior to t and $d(t)$ is the departure rate at t. The amount of time spent in the queue by a vehicle which arrives at the facility at time t ($T(t)$) is found using

$$T(t) = \frac{N(t)}{s} \qquad (1.12)$$

The total delay, $X(t)$, caused by the queue up to time t can also be calculated:

$$X(t) = \int_0^t a(u) - d(u)du \qquad (1.13)$$

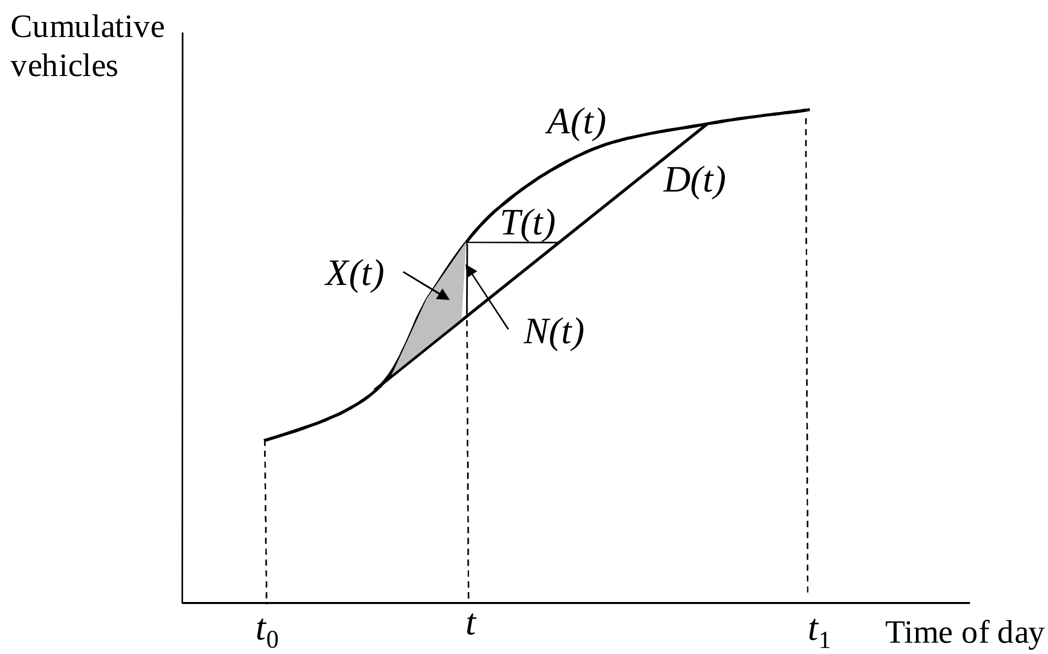


Figure 1.3: Simple queuing diagram

Queuing is analysed as a stochastic process when either the arrival rate or the service rate (or both) are random variables. A set of steady-state equations exist to describe stochastic queuing phenomena but they are of little utility for the modelling of individual vehicles. A related area of study deals with the strategies drivers employ to choose between queues. This type of analysis deals fundamentally with the effect of information on equilibrium. Situations often arise where multiple queues exist in parallel and drivers have the opportunity to choose the queue that will minimize the amount of time they spend waiting. A common example is the toll plaza. In such cases, queue jockeying (switching from one queue to another) can occur. Assuming drivers wish to minimize the amount of time they spend waiting in line, the length of each queue is the key piece of information that will determine queuing behaviour. In the case of an observable queue, its length can be known before the decision of whether or not to join must be made. In an unobservable queue, the length cannot be known until after the choice of queue has been made (Hassin & Haviv, 2003). Attributes of the queue, such as its expected length and expected duration, therefore depend upon whether or not the queue is observable.

1.3.1.3 Models of infrastructure

Both the car-following and queuing models are applied in the design and evaluation of road infrastructure. For example, the Highway Capacity Manual (HCM; Transportation Research Board, 2000) uses the car-following model for the analysis of uninterrupted flow facilities (freeways and highways) and the queuing model for at-grade intersections. On freeways and highways, the delay model (Figure 1.4) is based on the uncongested portion of the volume-delay relationship developed using the fundamental speed-flow-density relationships and validated with observations on the ground.

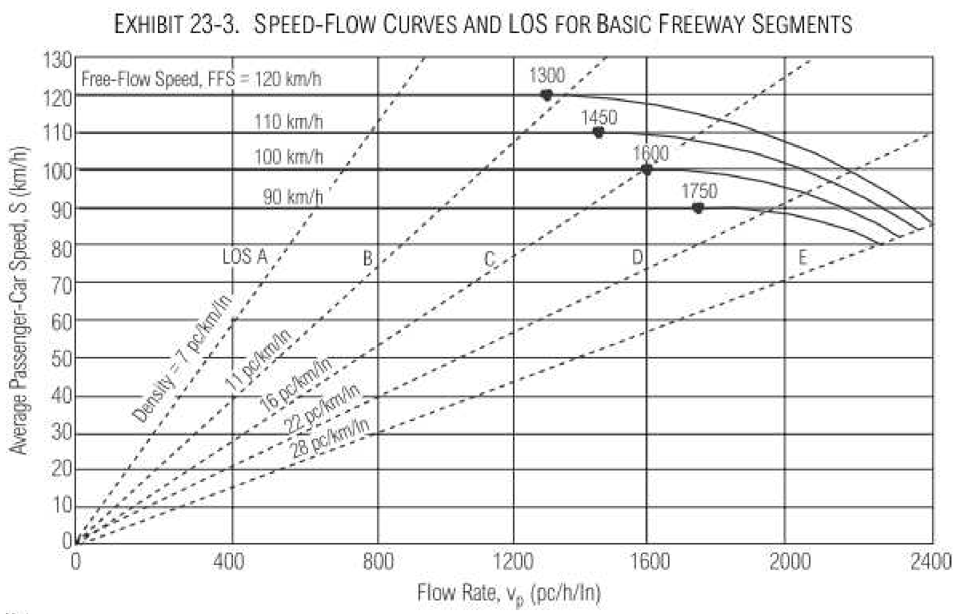


Figure 1.4: Speed-flow relationships for basic freeway segments (Transportation Research Board, 2000).

An important aspect of the HCM methodology is the road classification system. The HCM recognizes that different facilities may serve different purposes. A distinction is made, for example, between roads which provide maximal mobility versus roads which are designed to provide access (HCM Ch. 10). Intuitively, the objectives of accessibility and mobility are incompatible since the former favours high speed travel while the latter injects and withdraws vehicles from the traffic stream. The turning movements and parking manoeuvres which are features of access roads tend to disrupt the traffic flow and lead to a reduction in traffic speeds.

The Highway Capacity Manual measures road performance using a metric called level-of-service. The measurement system consists of the six letters from A to F. On freeways and other uninterrupted flow facilities, each letter represents a range of per-lane traffic densities. Density is an appropriate indicator since it dictates traffic speed, vehicle manoeuvrability and overall comfort of driving. Level-of-service on interrupted flow facilities depends on the delay imposed by the signalling regime. The HCM outlines detailed procedures for estimating delay caused by two-way, three-way and four-way stops, as well as for signalized intersections. These methods are based on a combination of queuing theory and empirical measurement. The localisation of traffic signals and design criteria are described in the Federal Highway Administration's Manual on Uniform Traffic Control Devices (MUTCD; U.S. Department of Transportation Federal Highway Administration, 2007). In particular, the manual sets out guidelines for the implementation of traffic signals based on observable quantities such as traffic volumes on each approach, pedestrian volumes and total delay. The idea is to maximize the efficiency of the urban street network while ensuring safety. The primary interest of this research is the aggregate effect of the control system on driver behaviour. Some of the basic principles of traffic signal control are examined in consequence.

According to the HCM, "a traffic signal essentially allocates time among conflicting traffic movements that seek to use the same space" (Transportation Research Board, 2000, p. 10-9). This task is accomplished through a repeating sequence of indications known as a cycle which has a specified period or length, C. Each movement (through, right, left, etc.) at an intersection corresponds to a phase, defined as a sequence of time intervals during which different indications are displayed. These indications use the universally understood colours red, yellow (amber) and green. A single phase can apply to multiple movements, and a single movement can be governed by more than one phase. Movements are made from lane groups which are defined according to the intersection geometry. Following the HCM convention, a lane group is represented by the letter i. In the majority of cases, no lane group has a green interval (G_i) which lasts for the entire cycle. The capacity, c_i, of a lane group is therefore a function of its corresponding green interval as a proportion of the total cycle length, C. The lane group capacity also depends on the total "lost time" (t_L) during each cycle. An illustration of the concept is shown in Figure 1.5. The lost time is the time required for vehicles to transition from a standing queue to the maximum flow rate known as the saturation flow rate (s). The perception-reaction time of drivers and the acceleration rate of vehicles determine the lost time. Typical default values for lost time are 3-4 seconds per phase. The saturation flow rate is the maximum rate at which vehicles can pass through the intersection. A commonly assumed value is around 1800 passenger cars per hour per lane (pcphpl) although it can vary greatly according to traffic stream composition, intersection geometry and localized phenomena such as on-street parking and bus stops. Since some time is required for traffic to stop at the end of the phase, the lost time is often incorporated into the yellow interval (Y_i). The total lost time per cycle, L, is the sum of the lost time for each critical phase (discussed below). Generally speaking, a larger number of independent movement starts (phases) produces a greater amount of lost time per cycle.

To summarize, the capacity of lane group *i* is computed as:

$$c_i = \frac{s g_i}{C} \qquad (1.14)$$

where

$$g_i = G_i + Y_i - t_L \qquad (1.15)$$

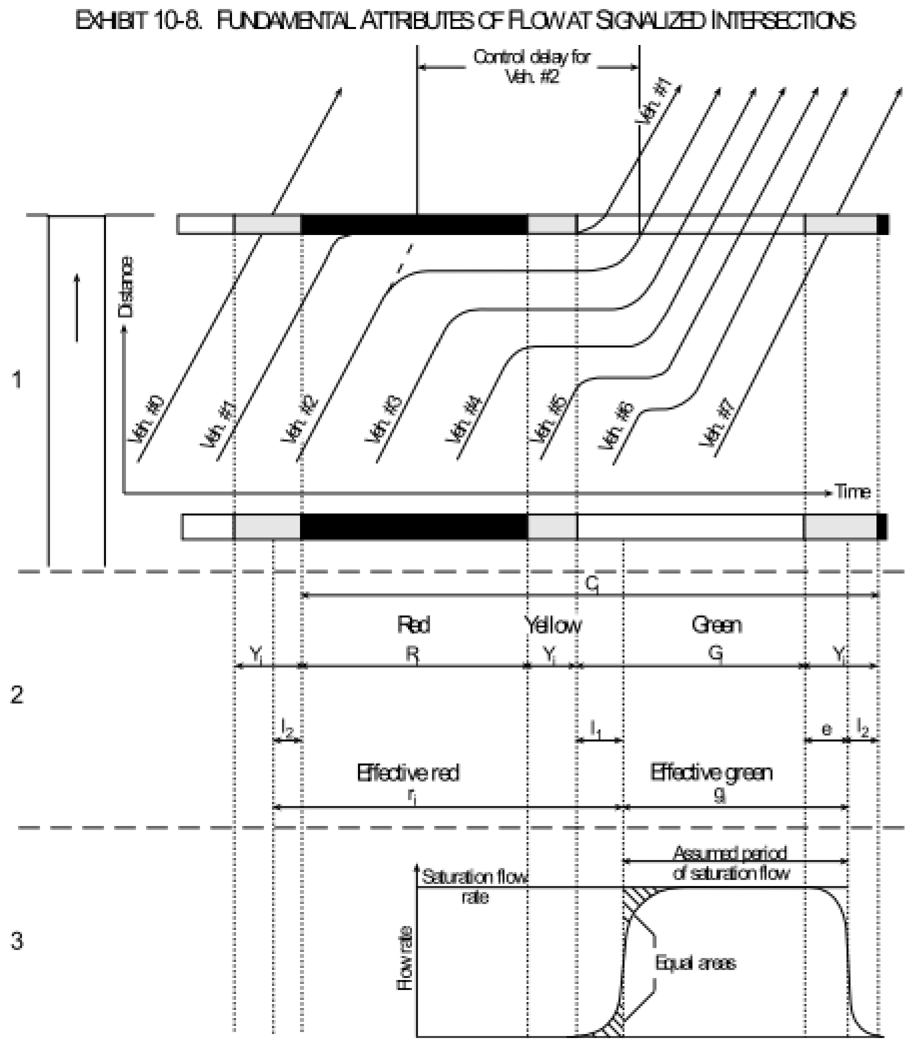


Figure 1.5: Detailed illustration of the phenomenon of lost time at a signalized intersection (Transportation Research Board, 2000).

The level-of-service of a signalized intersection or an approach to a signalized intersection is based on an evaluation of delay. The HCM describes total delay as “the difference between the travel time actually experienced and the reference travel time that would result during base conditions: in the absence of traffic control, geometric delay, any incidents, and any other vehicles” (p. 10-16). In principle, the delay at an unsaturated intersection for a single approach can be expressed simply as:

$$D = \frac{ar^2 s}{2(s-a)} \qquad (1.16)$$

where a is the rate of arrivals, r is the duration of the red interval and s is the saturation flow rate. In practice, signal synchronization and the lost time make the computation more complicated. In cases where incoming traffic is governed by an upstream traffic signal, vehicles tend to arrive in platoons. If the two traffic signals are properly synchronized, most of the platoon will clear the downstream intersection during the green interval. If the signals are not synchronized, a considerable portion of the platoon will reach the downstream intersection during the red interval and a queue will form as a result. This queue can force vehicles which arrive subsequently to slow down or stop. Meanwhile, a longer cycle length will result in a longer queue. If the cycle length is too short, however, the lost time will consume a greater percentage of the cycle and reduce the amount of effective green.

In principle, signalized intersections in urban areas are designed with a certain amount of synchronization although the network-wide optimization of traffic signals is a complex undertaking. In a steady-state analysis which does not consider synchronization, the expected delay is based on the saturation ratio, w, which is defined as the arrival flow rate a divided by the saturation flow rate s.

$$w_i = \frac{a_i}{s_i} \qquad (1.17)$$

This quantity is incorporated in Webster's formula for the cycle length (Webster & Cobbe, 1966) which minimizes the expected delay, given the opposing effects produced by queuing and lost time:

$$C_{opt} = \frac{1.5L + 5}{1 - W} \qquad (1.18)$$

where W is the sum of critical saturation ratios for the intersection. Generally, a critical saturation ratio is the highest saturation ratio among movements which share a phase.

While these methods inform design procedures, more precise calculations are performed through dynamic simulation of individual vehicles (microsimulation). A microsimulator nonetheless requires aggregate as input estimates of demand between origin-destination pairs. These estimates are often based on larger-scale simulations derived from the four-stage approach to regional transportation modelling.

1.3.2 The Four Stage Paradigm

While the Highway Capacity Manual is used to model the performance characteristics of road facilities, a separate methodology is required for estimating the demand for a particular facility. Demand for transport at a particular location is the result of a series of choices made by individual travellers. The goal of the four stage model is to predict these choices, especially in an urban environment.

In its purest form, the four stage model converts land use and population data into vehicle or person flows on network links using the following sequence of procedures: trip generation, trip distribution, mode split and traffic assignment (Figure 1.6). At its most elaborate, the four-stage model provides an integrated forecast of transportation costs and urban development. For example, the QRSII software (Horowitz, 2004) incorporates every aspect of the paradigm from land use forecasts to dynamic traffic assignment. The results of the assignment process can be used as feedback in the trip distribution step.

The initialization of the four-stage model is the division of the urban area into a set of analysis zones. All the calculations which follow use these zones as their primary unit. Over time, the modelling environments have evolved to incorporate dynamic variables, discrete choice methods, cellular automata and neural networks. The culmination of this evolution is activity-based models that incorporate land use (Habib, 2007; Mahmassani, 2006; E. Miller, Roorda, & Carrasco, 2005; Vovsha & Bradley, 2006). Despite the complexity of these emergent methods, they remain by and large fixed in the aggregate framework of the four-stage paradigm: in most cases their population of independent decision-making agents must be synthesized using data aggregated to the level of zones and the traffic simulations in particular retain numerous artificial constructs such as centroids, o-d matrices and volume-delay functions.

It is easy to criticize the four-stage model and, indeed, many authors have done so. Nonetheless, it remains widely used in practice because the vast majority of the commercially distributed tools (particularly simulation software packages) use it as a basis. Two specific shortcomings of the four-stage model are addressed here: the aggregation of information into an arbitrary system of zones and the lack of unique solutions.

The act of aggregating is very far from a trivial exercise. To begin with, it gives rise to the modifiable areal unit problem (MAUP) which means that the result of the analysis is entirely dependent upon the zone system definition. Moreover, the aggregation procedure can destroy a great deal of pertinent information. It does this in two ways: first, it replaces a distribution of values by a single mean and second, it permanently separates a transportation object from its attributes. An example of the latter phenomenon is the conversion of individual trips into an origin-destination matrix based on a system of zones. The trip, which is performed for a particular purpose at a particular time by a person of a certain age and gender, is reduced to a single numeric value representing the weight of the trip in the simulation. The theoretical objections to the use of an aggregate platform are accompanied by practical ones. For instance, it is virtually impossible to transfer information between independent platforms unless they have identical aggregation formulas.

The issue of solution uniqueness arises from methods of optimization. Many of the objective functions constructed in the urban transportation planning process cannot be resolved analytically. The search for minima and maxima almost always requires simulation using numerical methods. In such situations it is pertinent to ask whether the optimum (if it can be found) is a global optimum or a saddle point. It is also important to ascertain whether the solution generated by the algorithm is unique. If the objective function is almost flat, as is often the case near the optimal point, the optimal solution will have many near-equivalents.

The next four sections detail each of the steps in the four-stage model. The generation, distribution and mode split steps are described primarily in terms of the useful mathematical concepts upon which they are based. The traffic assignment step is discussed in greater detail since it is the focus of the present research.

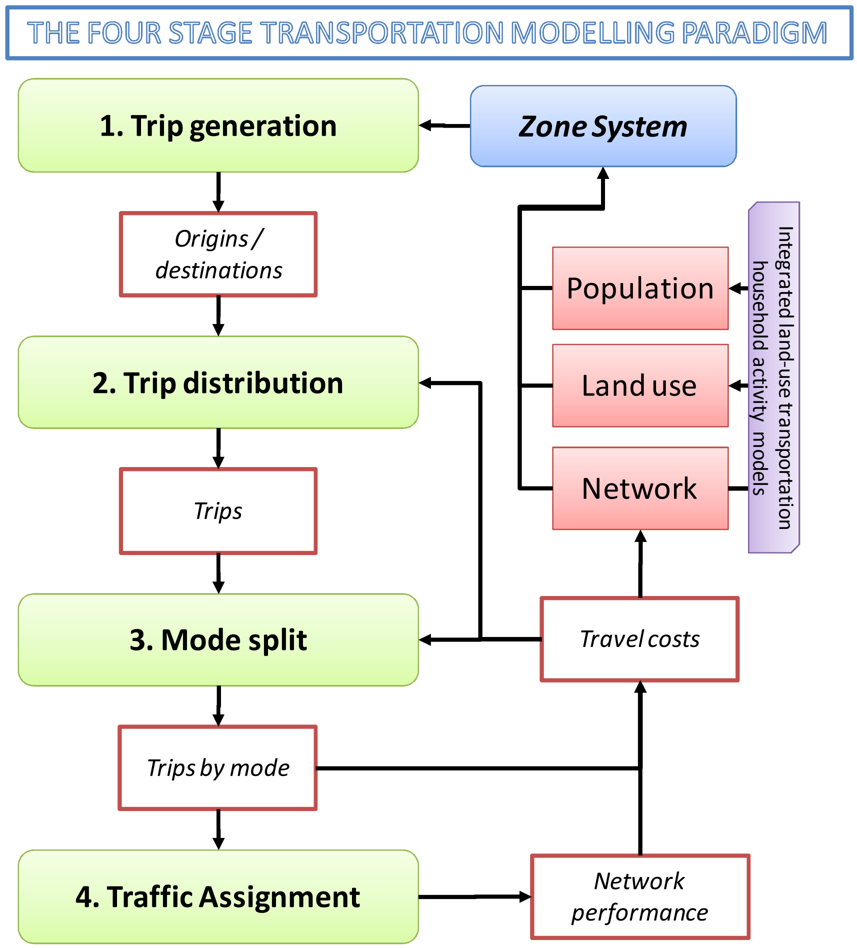


Figure 1.6: Diagram of the four-stage model or transportation planning

1.3.2.1 Trip generation

The trip generation step was devised to generate estimates of travel demand in situations where direct observation is impossible. The model takes as input information describing the spatial distribution of different economic activities and produces as output a number of trip ends (origins and destinations or attractions and productions) per zone. These data are often derived from a population census or a land use map. Trip generation models are typically based on linear regression methods where a dependent variable is assumed to be correlated with multiple observable quantities as well as unobservable ones. The underlying assumption is that unobserved factors also play a role but that their aggregate effect is unimportant. They are simply background “noise”. Stated more formally, an ordinary least-squares linear regression model will make predictions which differ from observations. The mean value of the square of these differences, however, will be zero.

In the case where the variation of an observable quantity y is linearly dependent upon the variation of other observable quantities x_0 through x_k as well as the behaviour of normally distributed unobservable effects ε, it is possible to construct a model of the following form:

$$y_i = x_0 b_0 + x_1 b_1 + \cdots + x_k b_k + \varepsilon \qquad (1.19)$$

Or, in matrix form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (1.20)$$

Predictions of **Y** can be obtained through the expression

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \qquad (1.21)$$

where

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad (1.22)$$

and **X** and **Y** are observed directly.

The authoritative reference on this topic is the ITE Trip Generation Handbook (Institute of Transportation Engineers, 2004) which provides estimated parameters for linear regression models of trip generation. Separate parameters are provided for different land uses. The result of the trip generation step consists of two vectors, one containing the number of trips produced (origins) in each zone and the other containing the number of trips attracted (destinations) to each zone.

The estimation of linear regression models is common practice in transportation planning. Care must be taken, however, to ensure that the estimated parameters are sensible and, just as importantly, that the observations used for the estimation are coherent. An assessment of the model parameters too often takes precedence over a thorough exploration of the data.

Assumptions which are fundamental to the approach go unverified as a result. These assumptions include a linear relationship between *x* and *y*, non-correlated independent variables, a normally distributed error term, and an error term distribution with constant variance.

1.3.2.2 Trip distribution

Once the number of trip attractions and productions is known for each zone, it is necessary to convert these trip ends into trips. Mathematically, it is the conversion of the two vectors produced by the trip generation step, each of dimension *n* (the number of zones), into an *n* by *n* matrix of trips. In the absence of more detailed information, a model of spatial interaction is constructed. The underlying hypothesis of the model is that the amount of travel between two locations is inversely proportional to the distance between them and directly proportional to their combined trip-generating capacity. For obvious reasons, these types of models are referred to as gravity models.

The estimation of the model parameters is typically performed by optimizing a non-linear objective function subject to a set of constraints. The objective function can represent the total cost of transportation, which must be minimized, or the total entropy, which must be maximized. A discussion and derivation of the numerous possible forms of such models can be found in Tobler (1988). An observed origin-destination matrix can be used in place of a travel cost matrix to perform the iterative balancing of trip ends.

The implications of applying a trip distribution model are not necessarily obvious. Despite its widespread adoption, most practitioners are unaware that the trip distribution algorithm involves the optimization of an objective function. Moreover, entropy maximization is appropriate for situations where only a minimal amount of information describing a particular phenomenon is known. While situations of severe information scarcity are still common, they are becoming less so. Existing data should be exploited to the greatest possible extent before the application of such a naïve model.

1.3.2.3 Mode choice

The choice of mode is typically presented as the third step in the 4-stage process but it is in some respects a deviation from the procedural framework. Aggregate models of mode choice are usually implemented in the two earlier stages. Trip generation rates for different modes are widely available and the gravity model can be applied using the cost of travel by any mode.

The bulk of the academic literature on mode choice deals with the method popularized by the Nobel laureate Daniel McFadden. Based on the microeconomic concept of utility maximization, this approach consists of a generalized linear regression model that is applied to individual actors (McFadden, 1974). In this respect, it is a disaggregate model incompatible with the zone system used in the generation and distribution steps. The conventional results of the model, however, are aggregate in the sense that they predict market share, rather than individual choice.

A summary of the algebra is as follows:

The utility of individual i is assumed to have observed and unobserved components V_i and ε_i, respectively. Total utility, U_i, is therefore

$$U_i = V_i + \varepsilon_i \qquad (1.23)$$

If the unobserved (random) component of the utility is assumed to follow a Gumbel distribution, then it is possible to estimate a generalized linear regression model where the dependent variable is the probability of person i choosing mode l. The form of the model is

$$\Pr(m_i = l) = \frac{e^{U_{l,i}}}{\sum_{j=1}^{k} e^{U_{j,i}}} \qquad (1.24)$$

$$\widehat{U}_{j,i} = \mathbf{X}_j\widehat{\boldsymbol{\beta}} + \mathbf{Y}_i\widehat{\boldsymbol{\gamma}} \qquad (1.25)$$

where m_i is the mode chosen by person i, $U_{l,i}$ is the utility person i obtains from mode l and k is the number of modes among which person i can choose. The utility function is a linear combination of variables representing attributes of mode j ($\mathbf{X}_j$) or attributes of person i ($\mathbf{Y}_i$). The parameters $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\gamma}}$ are estimated using maximum likelihood techniques.

The logit model has the property of independence of irrelevant alternatives (IIA) which means that the attractiveness of one option relative to another does not depend upon the total number of options. The flip side of this property is that absolute market share of each option *does* depend on the number of options. It is essential, therefore, that the choice set offered to each decision-maker be composed of truly independent alternatives. There are many situations where the choices are not independent, such as the choice between auto-drive and park-and-ride modes. Considerable effort has been expended to circumvent the IIA property through the application of other methods such as the probit model, the nested logit model (Ortuzar, 1983) and the mixed logit model (Hess, Bierlaire, & Polak, 2005; McFadden & Train, 2000). These same methods have also been applied to models of route choice, discussed in more detail below.

As is the case with trip generation and trip distribution methods, it is important that the mode choice model estimation results not take precedence over trends which are evident in the data. For example, a notorious property of logit models (and regression models in general) is the possibility of estimating statistically significant parameters with incorrect signs. A more fundamental problem with utility maximization theory arises from the demands it makes of individual travellers. In particular, it assumes that travellers have perfect information about the alternatives available to them and that they can accurately predict the consequences of each choice. These assumptions are necessary for the algorithm to function, but they are not always representative of reality.

1.3.2.4 Traffic assignment

Traffic assignment – the last step in the 4-stage model – is the step of greatest interest to the civil engineer since it estimates the load on the built infrastructure and permits an evaluation of the performance of the system. The load on the built infrastructure is represented by the vehicle flows on network links and vehicle flows are the aggregation of the paths followed by individual travellers. Traditionally, a traffic assignment model takes the origin-destination matrix for each mode that is typically generated using the previous 3 steps and assigns it to the network using a rule for route selection.

The problem can be viewed from at least two perspectives. First, the optimal assignment of vehicles to a network is a very old logistical problem and is of interest to any transportation services provider, be it the military, a railway, an airline or a trucking company. These organizations engage in transport planning in the truest sense of the term: all movements are planned to the greatest extent possible. Second, in the particular case where a central authority provides only the right-of-way, but not the vehicles, the situation is much more chaotic and the relevance of the term "planning" is not always obvious. Nevertheless, techniques have been designed to predict the distribution of independent vehicles on a partially controlled network. These methods are applied to the modelling of urban automobile traffic. The three components of a traffic assignment model are: a representation of transport supply, a representation of demand, and a hypothesis about driver behaviour. In the next three sections these elements are described in detail.

1.3.2.4.1 Representation of Transport Supply

Transport supply is conventionally represented using the concepts of links and nodes, or arcs and nodes, or edges and vertices, borrowed from graph theory where a graph, *G*, is composed of a set of links, *L*, and nodes, *N*.

$$G \coloneqq (L, N) \tag{1.26}$$

In general, the links in a graph representing a transportation network are directed meaning that each link *l* is defined by an ordered pair of nodes.

$$l = (x, y) \tag{1.27}$$

where *x* and *y* represent, respectively, the “head” and “tail” nodes of the link.

The graph-based model seems appropriate since terrestrial transport networks, when viewed from the air or on a map, appear as a series of intersecting lines. The visual appearance of the system dominates the construction of the graph theory isomorphism, with intersections represented as nodes and the road segments between intersections as links. The origin and destination of a trip are usually taken to be nodes, although alternative approaches do exist (Federal Highway Administration, 2005; Horowitz, 2001).

A fundamental tenet of transportation modelling is that travel between two points has a cost. It is usually assumed that travel costs associated with a trip are incurred on links rather than at the nodes. In static traffic assignment models, it is common practice to assign penalties and prohibitions to movements between certain links. These transfer penalties can be used to represent legal restrictions on intersection movements or to represent the additional delay associated with a particular type of movement, such as a left turn. The advent of dynamic microsimulation has led to the explicit representation of queues at intersections and the delay caused by waiting for gaps in opposing traffic, thus obviating the need for the application of movement-type penalties.

In computerized models of transport systems, the network is structured as a set of related databases. The database of nodes, which represents intersections and serves to define all the other objects in the network, indicates the location of each node in space and contains relevant information about the type of traffic control which is implemented. The direction of each link is defined based on the order of the two nodes to which it is connected. Other link attributes include travel time, capacity and functional class. Another database describes which movements (m) are possible between links at a given intersection (node i). Formally,

$$m(i) = (r, s) \qquad (1.28)$$

where

$$r = (h, i) \qquad (1.29)$$

$$s = (i, j) \qquad (1.30)$$

In other words, r is a link which enters node i and s is a link which exits node i. There are numerous types of movements, the most common being through, right turn and left turn. Other varieties include u-turns and merges. The type of movement is defined by the angle between the entering and exiting link. A simple example of a typical model intersection is shown in Figure 1.7.

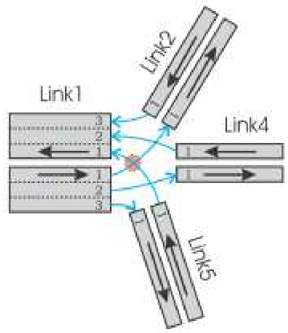


Figure 3-3: Lane Connectivity (blue arrows) for the lanes from links related to Node 2.

Figure 1.7: An example of nodes, links, lanes and movements constructed for traffic modelling purposes (Serras, 2007).

At unsignalized intersections, each movement is assigned a priority according to the position of stop or yield signs. At signalized intersections, movements are associated with a signal phase and each phase belongs to a cycle. All these objects are explicitly represented in dynamic traffic assignment models, but are necessarily ignored in static models.

In urban road networks the flow of traffic is regulated primarily by control systems at intersections. In addition, path choice sets are defined by movement permissions and capacities. Although links remain associated with a cost of travel, they serve primarily to connect intersection movements. This conception of the road network lends itself to the construction of a “dual” problem, where road segments become nodes and intersections are exploded into links representing permitted movements (Añez, de la Barra, & Pérez, 1996; Hu, Jiang, Wu, Wang, & Wu, 2008; Porta, Crucitti, & Latora, 2006; Volchenkov & Blanchard, 2007).

The attributes of links, nodes and movements are defined according to a hierarchy of functional classes. The concept of hierarchy is well-established and formally recognized by modellers in the public transit domain but is usually much less explicit within the traditional road modelling paradigm (for a detailed discussion, see van Nes, 2002). There are many different ways of classifying road network elements. Some common designations applied to network links are: freeway, arterial, collector, local and ramp. Chapter 5 of the HCM provides general definitions for each of these categories except local roads. These definitions are reproduced below.

Freeway – A multilane divided highway with a minimum of two lanes for the exclusive use of traffic in each direction and full control of access without traffic interruption.

Arterial – A signalized street that primarily serves through-traffic and that secondarily provides access to abutting properties, with signal spacings of 3.0 km or less.

Collector – A surface street providing land access and traffic circulation within residential, commercial and industrial areas.

Ramp – A short section of roadway connecting two traffic facilities.

Integral to these definitions are the notions of mobility and accessibility. Freeways and arterials primarily serve through traffic in order to maximize mobility. Collector roads and ramps are designed to provide access, the former by connecting to different land uses and the latter by linking one road facility to another. Intuitively, local roads also are designed for access rather than mobility purposes.

A scientific method for applying this classification system remains largely elusive, although chapter 10 of the HCM does discuss in some detail the importance of a functional hierarchy for the maximization of accessibility and mobility. The HCM functional and design criteria for distinguishing between different road classes are shown in Figure 1.8. The degree to which these criteria are applied in the official classification of roads by local authorities varies significantly between jurisdictions. For the purposes of simulating the route choices of drivers, however, the official classification matters less than the apparent functional class of the road.

EXHIBIT 10-4. FUNCTIONAL AND DESIGN CATEGORIES

	Functional Category			
Criterion	Principal Arterial		Minor Arterial	
Mobility function	Very important		Important	
Access function	Very minor		Substantial	
Points connected	Freeways, important activity centers, major traffic generators		Principal arterials	
Predominant trips served	Relatively long trips between major points and through-trips entering, leaving, and passing through the city		Trips of moderate length within relatively small geographical areas	
	Design Category			
Criterion	High-Speed	Suburban	Intermediate	Urban
Driveway/access density	Very low density	Low density	Moderate density	High density
Arterial type	Multilane divided; undivided or two-lane with shoulders	Multilane divided; undivided or two-lane with shoulders	Multilane divided or undivided; one-way, two-lane	Undivided one-way, two-way, two or more lanes
Parking	No	No	Some	Significant
Separate left-turn lanes	Yes	Yes	Usually	Some
Signals/km	0.3–1.2	0.6–3.0	2–6	4–8
Speed limit	75–90 km/h	65–75 km/h	50–65 km/h	40–55 km/h
Pedestrian activity	Very little	Little	Some	Usually
Roadside development	Low density	Low to medium density	Medium to moderate density	High density

Figure 1.8: Functional and design categories of urban streets according to the 2000 Highway Capacity Manual (Transportation Research Board, 2000).

In many traffic models, the hierarchy is implied by the capacity and speed attributes of each segment. Under congested conditions, however, these characteristics vary with demand and so the functional hierarchy fades as flow levels increase. In other words, the freeway becomes as slow as the adjacent arterial road. The hierarchy may nevertheless remain in the user's perception of the network. It has been suggested that the hierarchy is an "emergent" property of networks (Yerra & Levinson, 2005), meaning that certain roads will collect disproportionate shares of traffic even if no conscious effort is made by an overseeing authority to establish different road types. This tendency may be related to the greater connectivity of certain roads relative to others (Lämmer, Gehlsen, & Helbing, 2006).

Until quite recently, the possibility of modelling the entire road network of a large city seemed remote due to constraints on computer memory and processing power, In addition, the aggregation of demand into zones means that a trip origin and destination are represented by a single point (the zone centroid) connected to the network by artificial links. The introduction of such a large spatial distortion renders unnecessary the codification of a detailed network. Centroid connectors are considered proxies for local roads and are usually coded as uncongested, low-speed facilities as a result. Although current technology is sufficiently powerful to make these contrivances unnecessary, the limited amounts of reliable demand data in most jurisdictions means that a system of zones must be retained.

1.3.2.4.2 Representation of Transport Demand

The aggregate nature of the 4-stage model means that demand is input into the assignment procedure in the form of one or more origin-destination matrices, each representing a time period or “slice”. In large-scale regional models, each slice typically represents at least one hour. In more microscopic models, the time slices are often in minutes. Algebraic convenience aside, the use of o-d matrices has little to recommend to it. In early models, where zones were small in number and large in size, the use of a matrix was justifiable especially considering the formerly high cost of computer memory. Over the years, there has been a trend toward smaller zones which means that more of them are required to cover the same area. A smaller zone will generate less total demand and this demand will be distributed over a larger number of origins and destinations. As a result, it is common to find matrices composed of millions of cells with the vast majority (95% or more) empty or containing a microscopic number of trips. In practice, it is common to find non-zero o-d pairs with demand less than 1. In such cases, the matrix becomes an extremely inefficient way to store trip information.

Another problem with the o-d matrix is its diagonal elements. Demand along the diagonal is demand within the zone and, since all demand for a zone originates or terminates at a unique centroid, such trips are not assigned to the network. In zones which cover a large geographical area, the number of trips not included in the assignment may be significant.

The greatest criticism that can be made of the o-d matrix is the paucity of information that it contains. Travel demand is in reality a complex phenomenon dependent upon attributes of individuals, households, vehicles and activity schedules. An o-d matrix, as a two-dimensional table, necessarily ignores socio-demographics, household interactions, and the evolution of demand over time. While matrices can be constructed for small time periods, it is absurd to employ, for example, matrices for each minute in a 3-hour simulation particularly given the data storage issues discussed above. The logical solution to these problems is to input demand in the form of a list which, in addition to being more efficient, would also allow for the assignment of relevant traveller attributes to the network. The TRANSIMS package (Federal Highway Administration, 2005), designed for activity-based modelling, offers this possibility.

1.3.2.4.3 Representing Supply-Demand Interactions (Equilibrium)

A model of demand-supply interactions for automobile traffic requires a hypothesis about the way drivers choose their routes. The simplest hypothesis is one which says that drivers choose the path which offers the minimum travel time between their point of origin and their destination. When all the demand for a particular o-d pair is assigned to the shortest path (Bellman, 1958; Dijkstra, 1959) using this hypothesis, the model is called an all-or-nothing assignment. The major weakness of this approach is that it does not account for congestion effects.

Traffic assignment models which account for congestion were born out of Wardrop’s hypotheses (principles) concerning the behaviour of drivers. The first hypothesis says that:

“The journey time on all the routes actually used are equal, and less than those which would be experienced by a single vehicle on any unused route.”

This hypothesis is known as the user-equilibrium principle. The second principle, dubbed the system equilibrium, says simply:

“The average journey time is a minimum” (Wardrop, 1952).

Despite the difficulties involved in obtaining supporting empirical, the very intuitive first principle is applied almost universally in models of congested traffic. The assumption of user-equilibrium in a transport network has several important implications. Most obviously, the total cost of travel is not necessarily minimal, as would be the case under a system-equilibrium. The cost of travel between a given o-d pair is not necessarily minimal either, even though no user can reduce his travel time by unilaterally switching paths. More surprisingly, it is possible that the removal of certain links can actually reduce the user-equilibrium travel time between a given o-d pair. This last phenomenon is known as Braess’ paradox (Braess, Nagurney, & Walkobinger, 2005).

Wardrop's principles were expressed mathematically by (Beckmann, McGuire, & Winsten, 1956) and a functional algorithm for solving the problem was developed by (Frank & Wolfe, 1956). A summary of the procedure is given below, as described by (Sheffi, 1985).

The model is based on the assumption that there exists, for each link *a*, a relationship between the traffic flow (*f*) and the travel time (*t*). Letters in bold represent vector quantities. This relationship is known as a volume-delay function. Under the hypothesis of user-equilibrium, the objective function (*z*) to be minimized is

$$z(\mathbf{f}) = \sum_a \int_0^{f_a} t_a(u)du \qquad (1.31)$$

subject to constraints on path flow:

$$\Phi_{rs} = \sum_k f_{rs,k} \qquad (1.32)$$

$$f_{rs,k} \geq 0 \qquad (1.33)$$

where f_a is the flow on link *a*, t_a is the travel time on link *a*, Φ_{rs} is the total demand between origin *r* and destination *s* and *k* is the path index. The variable $f_{rs,k}$ is therefore the flow on path *k* between origin *r* and destination *s*. The objective function *z* has no real-world analog. Link flows are related to path flows through the expression

$$f_a = \sum_\Omega \sum_k f_{\Omega,k} \delta^a_{\Omega,k} \text{ where } \delta^a_{\Omega,k} = \begin{cases} 1 \text{ if link } a \text{ is part of path } \Omega,k \\ 0 \text{ otherwise} \end{cases} \qquad (1.34)$$

where Ω is the o-d pair index.

The vector of all link flows is represented by **f**. Since the travel time on a given link is dependent on the amount of flow using the link, the optimization process is necessarily iterative. At each iteration *n*, a new value of the objective function is found by performing an all-or-nothing assignment. This process yields an auxiliary set of link flows **g**. The objective function at iteration *n* becomes

$$z^n = \sum_a t_a \, g_a \tag{1.35}$$

The amount of flow to transfer between regimes **f** and **g** is computed by taking the derivative of the objective function evaluated at some step size α along the descent direction ($\mathbf{g}_n - \mathbf{f}_n$), setting it equal to zero and solving for α.

$$\frac{\partial}{\partial \alpha} z[\mathbf{f} + \alpha(\mathbf{g} - \mathbf{f})] = \sum_a (g_a - f_a) t_a [f_a + \alpha(g_a - f_a)] \equiv 0 \tag{1.36}$$

To proceed to the next iteration, set

$$\mathbf{f}^{n+1} = \mathbf{f}^n + \alpha(\mathbf{g}^n - \mathbf{f}^n) \tag{1.37}$$

The algorithm terminates when the link flow distribution is sufficiently close to the theoretical user-equilibrium state. A common metric for representing the disparity is the relative gap, R^n, which is the difference between the total cost of travel using the link flow assignment of the current iteration and the total cost of travel if all flow were assigned to the shortest path, divided by the current value of the objective function. In other words:

$$R^n = \frac{\sum_a t_a(f_a^n)(f_a^n - g_a^n)}{\sum_a \int_0^{f_a^n} t_a(u) du} \tag{1.38}$$

Because the objective function does not represent a real physical quantity, the relative gap has no obvious interpretation, so alternative measures of convergence are often employed. The example below compares travel times between o-d pairs over successive iterations:

$$\epsilon > \sum_{rs} \frac{|u_{rs}^n - u_{rs}^{n-1}|}{u_{rs}^n} \tag{1.39}$$

where u_{rs}^n is the minimum travel cost between r and s at iteration n and ϵ is an arbitrarily small quantity.

Theoretical convergence of the algorithm can be guaranteed only if the link volume-delay function is monotonic increasing. Otherwise, the marginal cost of travel on a link could decrease or remain constant with increasing flow, thus permitting the assignment of an unlimited number of vehicles without increasing the value of the objective function. There exist numerous varieties of volume-delay functions (see for example Cheah, Dalton, & Hariri, 1992; Speiss, 1997) but the most common one is the equation proposed by the Bureau of Public Roads (Bureau of Public Roads, 1964).

$$t(v) = T_0 \left[1 + \alpha \left(\frac{v}{C} \right)^{\beta} \right] \quad (1.40)$$

where T_0 is the travel time on the link when the volume, v, is zero, C is the capacity of the link and α and β are calibration parameters. In practice, C and T_0 are often calibration parameters as well since their true values are difficult to obtain for every network link.

It is important to note that the solution at the end of each iteration is found in terms of links flows only. The existence of a theoretical unique solution can be proven mathematically. The path flow solutions, however, are non-unique. This property means that it is impossible to follow an individual vehicle through the network. In addition, the amount of flow on each path is not retained in the calculation process. As a result, the user-equilibrium method cannot properly be classified as model of route choice. It is actually a flow-optimization model. Simpler heuristics exist for optimizing flow on a network subject to capacity constraints. The Hitchcock method (Hitchcock, 1941) and the Ford-Fulkerson algorithm (Ford & Fulkerson, 1956) are two examples. These algorithms, however, do not incorporate the presumed selfish behaviour of independent drivers.

Because of its ability to represent congestion effects and Wardrop's user equilibrium principle, the Frank-Wolfe method (often referred to as deterministic user-equilibrium (DUE) traffic assignment) was successfully packaged and sold as commercial planning software beginning in the late 1970s (Achim & Florian, 1979; Florian et al., 1979). Competing approaches such as UTPS (Dial, 1971, 1976) were largely cast aside in favour of the DUE paradigm. Since then, simulations have increased in complexity and power but the underlying philosophy has remained fundamentally unchanged. Concerns have periodically been raised about the stability of the DUE solution, and as a result new algorithms have been developed which retain at least some path information and permit a better and more rapid convergence (Bar-Gera, 1999; Dial, 2006; Florian, Constantin, & Florian, 2009; Jayakrishnan, Tsai, Prashker, & Rajadhyaksha, 1994).

The weaknesses inherent to the approach are well-documented but often unappreciated by practitioners and, not infrequently, by researchers. When DUE models first became operational, computer memory was extremely expensive and as a result only skeletal representations of road networks were possible. In a large urban area, a skeletal network consisted primarily of freeways, which are uninterrupted flow facilities. On such roads, the average speed of traffic is deemed to be influenced uniquely by the level of demand. Average speed decreases as demand increases, as represented in the monotonic increasing volume-delay function. The applicability of this model to urban streets, however, is questionable since average traffic speed is primarily influenced by signal programming. There exists a body of research which demonstrates that the relationship between demand and delay on signalized corridors is considerably more complex, particularly if the signal system is responsive to traffic conditions (Gartner & Wagner, 2004; Lee & Machemehl, 2005; Meneguzzer, 1995). Moreover, the DUE model requires that a volume-delay function be specified for each link in the network. For large networks this task is considerable and, more importantly, the fidelity of the specification to any reality on the ground is difficult to verify in any systematic way.

A well-recognized shortcoming of the DUE model is that it is static. As a result, it cannot properly model dynamic phenomena such as queues at traffic signals and other choke points. A report from INRO (Mahut, Florian, Florian, Velan, & Tremblay, 2005) succinctly makes the point:

"[T]ransportation planners around the world are increasingly seeking traffic modelling tools that can account for the temporal effects of congestion. Static models are not designed to do this. The formation of queues and their eventual spill back to upstream links cannot be modelled appropriately, and flows that result on heavily congested links may be above capacity, which is not realistic."

As a result of this shortcoming, traffic assignment methodologies have fractured into three distinct approaches: macroscopic, mesoscopic and microscopic. The macrcoscopic approach has traditionally been applied in regional models where, due to computer memory constraints, networks, driver behaviour and simulation results are greatly simplified to produce a general portrait of conditions on a large scale. While useful for predicting demand for certain road facilities in terms of vehicular volumes, they are generally too coarse to reliably predict, for example, the expected number of turning movements at an intersection. Such information is deemed essential in the design of signalized intersections and the associated coordination schemes. The microscopic approach, which represents individual vehicles dynamically, was developed to meet this need. These models, however, must usually be fed traffic volumes as input since they usually do not simulate route choice. Mesoscopic simulators were developed to act as a bridge between the microscopic and macroscopic models (Barceló & Casas, 2006; Mahut et al., 2005). In this framework, individual vehicles are modelled dynamically and a route choice mechanism is incorporated to permit the assignment of traffic using origin-destination pairs as input. Unlike microscopic models, mesoscopic models are not based on discrete time intervals but rather on vehicle events, specifically the arrival and departure from a network node. This means that vehicle movement on links is not explicitly represented. Other dynamic traffic assignment (DTA) platforms have been developed, primarily for real-time testing of traffic management strategies implemented with the help of intelligent transportation systems (ITS) (Ben-Akiva, Bierlaire, Koutsopoulos, & Mishalani, 1998; Mahmassani, 2001; Peeta & Ziliaskopoulos, 2001). The ability of these platforms to represent entire urban areas was uncertain at the time of writing.

A large body of work adopts a distinctly different method to the problem of routing vehicles through a network. This route choice analysis approach uses the discrete choice theory commonly employed to estimate modal shares (Cascetta, Russo, Viola, & Vitetta, 2002; Frejinger, 2008; Hoogendoorn-Lanser & Bovy, 2007; Ramming, 2002). It is a probabilistic method which bears some similarity to the path-choice model developed by Dial (1971). The route choice and traffic assignment paradigms usually appear independently in the literature but some attempts have been made to unify the two theories by incorporating the equilibrium objective into the choice model (Bekhor & Reznikova, 2007; Chou, Takriti, & Underwood, 1993). The development of mesoscopic dynamic simulators and path-based traffic assignment algorithms (see above) represents a further step in this direction.

The collection of large quantities of complete path data is made possible through the instrumentation of vehicles with GPS. The generated data, which consist of a chronological sequence of points in space, can be matched with a digital network to produce detailed itineraries for a particular trip. In-depth analyses of this information structure reveal that driver route choice depends on numerous factors in addition to travel time. In particular, traffic signals, route directness, the time spent on the superior network and personal habit have been found to play an important role in the decision process (Bierlaire & Frejinger, 2008; Jan, Horowitz, & Peng, 2000; Papinski, Scott, & Doherty, 2009).

1.3.2.4.4 Equilibrium Re-examined

The analysis of disaggregate revealed preference route data undermines, to a certain extent, the user-equilibrium hypothesis. Jan et al (2000) note that:

"The current methods used by planners for modeling path choice in traffic assignment have been developed largely in the absence of objective empirical evidence of actual path choices. Theories of user-optimal equilibrium assignment and stochastic multipath traffic assignment have proven quite useful to planners, but those algorithms' underlying assumptions related to path choice have not received an adequate level of validation."

Moreover, the same study indicated that, in many cases, the chosen path was not the shortest. This finding is consistent with the hypothesis advanced by Scarlett who studied the path choices of drivers in Montreal during a snowstorm (Scarlett, 1970). Citing Simon (Simon, 1957), he points out that humans are more accurately described as "satisficers" rather than optimizers since they do not posses complete information about the available alternatives and because they are unlikely to follow an arduous computation process when making a decision. Moreover, it is clear that drivers cannot know with any certainty whether or not the path they choose is actually the shortest one at the moment they undertake their trip. They can choose their path only based upon their previous experience. To quote Scarlett directly:

"Drivers do not all, and repeatedly, experiment with alternate paths. Once settled on a satisfactory path, it is easier and safer to stay with it than look for an optimum. Habit is compulsive. But paths change over time in their optimality: growing congestion slows down a once-fast route... To the extent that he refrains from experimentation on alternate routes it may be weeks, months or even years before he discovers better alternate ways."

This is not to say that the user-equilibrium principle should be discarded since drivers do have a tendency to minimize the amount of time they dedicate to a particular trip. Nevertheless, two points must be emphasized. First, driver knowledge of network conditions at a given moment in time is necessarily incomplete and imperfect. Second, the tendency to minimize travel time is likely just that – a tendency. The degree to which travel-time minimization is imperative depends upon the individual, the trip purpose, the daily activity schedule and external factors such as non-recurrent congestion caused by traffic accidents.

1.3.2.4.5 Wardrop's Third Principle

The case for a more nuanced position with respect to the user-equilibrium hypothesis is provided by Holden (1989) who noticed a supplementary description of equilibrium in Wardrop's text and dubbed the sentence "Wardrop's third principle":

"Traffic will tend to settle down into an equilibrium situation in which no driver can reduce his journey time by choosing a new route."

Unlike the first two principles, this one refers explicitly to individual drivers, uses the word "equilibrium" and accounts for the probabilistic nature of the system through the verb "to tend". In addition, Holden points out that a transport network exists not so much in equilibrium as in a "steady state". Each day, forces are exerted which disrupt this state but the system always tends to return toward its "inbuilt tendency". The existence of a steady state may not be provable mathematically. But the fact that day-to-day travel time fluctuations are small enough to permit a rational choice of route of the type made by millions of drivers every day is strong evidence in its support. Furthermore, the steady-state hypothesis does not depend upon drivers choosing the shortest route, since even random route choices would tend toward a steady state.

Holden describes a road network in terms of a "state space" which is defined as:

"the set of all possible route-choice selections which do not exceed the capacity on any link and which are compatible with fixed OD demand."

Two attributes of the state space are the total excess time and the total travel time. Excess time is defined "for each driver as the time by which the route chosen exceeded the minimum possible journey time on that particular occasion." Clearly, the total excess time cannot be less than zero and the total travel time has a lower bound. A state in which the total excess time is zero constitutes a Beckmann equilibrium and the minimum total travel time corresponds to Wardrop's principle of system equilibrium.

Wardrop's third principle is an expression of the notion that the state of auto traffic on a network is the result of many route choices made by individual drivers. This state is quite different from the one implied by the word "assignment", where some higher power dictates the traffic levels on each link so that an aggregate measure of network performance conforms to a specified criterion.

1.3.2.4.6 Validation procedures

An elementary test of any model is the degree to which it can reproduce an observed reality. If it cannot do this satisfactorily then it cannot be relied upon to make meaningful predictions about projected scenarios. In the case of traffic models, "reality" is usually represented by roadside counts. The most common indicator for measuring the performance of a model is the amount of correlation between the observed vehicle volumes and the volumes forecast by the model. Very often, the roadside counts are performed by hand and for a single day only. Both the representativeness and reliability of the information are easily questioned as a result. More rarely, average traffic speeds are measured and then compared to the simulated speeds.

It can be argued that neither of these methods is appropriate for validating a model which purports to simulate path choices. A comparison of modeled paths with observed paths is required if the model is to be considered scientifically rigorous. Historically, observations of paths chosen by drivers have been very difficult to obtain. The fact that the models generate non-unique path-level solutions does not help matters.

1.3.2.4.7 Model calibration

The traditional approach to the calibration of traffic assignment models involves a great deal of ad-hoc manipulation. At its best, this exercise is based on certain ground truths. A good example is the discovery and elimination of network coding errors or inaccuracies. Using aerial photos or even more penetrating technologies like Google Streetview, it is a simple if often tedious matter to verify and correct the number of lanes, the speed limit, the presence of a traffic signal, turning restrictions and so forth. A more troubling example is the coding of zone centroid connector links which is necessarily an artful manipulation. Often, the discrepancy between simulated and observed traffic patterns cannot be rectified through supply-side interventions alone. On these occasions, it is not uncommon to adjust the demand by manually adding or removing trips in the origin-destination matrix. Since these manipulations are not based on any observed phenomenon, it is difficult to justify them as part of a scientific methodology. Moreover, translating these manipulations into projections of future demand is impossible.

While this calibration process often leads to improved accordance with reality (particularly when it comes to matching simulated volumes with traffic counts), it is very time consuming and the benefits are frequently marginal. Although network models should reflect reality to the greatest possible extent, arbitrary adjustments to representations of observed supply and demand must be minimized if the analysis is to retain any credibility as a scientific exercise

1.3.2.5 Critiques of the Four Stage Model

The four stage model has been widely criticized for its simplistic assumptions and lack of behavioural basis. At least two critics have gone further, attacking not just the four stage model but the urban transportation planning culture in general. Talvitie (1997, 2007) has argued that transportation planning, in its current form, is wholly unscientific and has failed to solve the problems it was originally designed to address. Moreover, an over-zealous belief in the principle of utility maximization and the equilibrium between supply and demand has been a contributing factor in many of the problems now faced by cities around the world: congestion, pollution, urban sprawl and economic segregation.

Several decades have elapsed since Atkins (1977, 1986) questioned the pertinence of urban transportation planning as embodied by the four-stage paradigm. Nonetheless, the points he raised remain highly relevant. To begin with, urban transportation planning was and is a big business. Governments all over the world give multi-million dollar contracts to private firms to “plan” transportation strategies. The usefulness and relevance of the end result are rarely questioned. Secondly, the plan as developed in the study is hardly ever executed due to widespread public opposition which was never foreseen by the planners. Third, the complexity of the modelling process makes it completely incomprehensible to anyone not intimately involved. Extreme complexity has two results: errors can propagate through the model undetected and an informed critique of the modelling procedure is almost impossible. Finally, urban transportation planning was conceived in an age when centralized planning by government was still publicly acceptable. Grandiose plans which aim to have a significant impact on people’s daily lives (and which will be a burden on taxpayers) are now viewed with deep suspicion.

These criticisms are not easily refuted. Performing a strategic analysis of public transportation infrastructure which conforms to basic professional and scientific criteria is a formidable challenge.

1.3.3 The Totally Disaggregate Information-based Approach

Developed at the École Polytechnique of the University of Montreal under the acronym MADITUC (Modèle d'analyse désagrégée des itinéraires de transport urbain collectif – Chapleau, Allard, & Canova, 1982; Chapleau, 1992), this approach is significantly different from the four-stage paradigm. It owes its existence to the presence of detailed information describing individual travel behaviour, including traveler interaction with the network obtained through descriptions of public transit itineraries (bus routes, stops and transfer nodes). This information is embodied in the Montreal household travel survey described in the next section. A reverse-engineered algorithm is employed to develop models which contribute to the mutual enrichment of the attributes of multiple transport objects. The platform is designed to optimize the operation of public transit systems and its functioning is based upon the direct observation of travel demand expressed in units of individual travellers. This information is typically collected using a revealed preference survey (described in the next section) although new perception technologies, particularly smart cards, offer potentially interesting alternatives. Assuming that sampling issues can be adequately dealt with, the information on chosen paths renders a considerable portion of the four-stage method obsolete since the quantity and distribution of trips, as well as the modal shares, are observed directly. In addition, partial information describing the utilized path is collected for many trips. The totally disaggregate nature of the data allows for post-treatment aggregation to any level of analysis, be it a zone, a transit line, a bus stop or a bridge.

A schematic example of the totally disaggregate information-based approach is shown in Figure 1.9. A transportation corridor is analyzed by constructing relationships between four independent sets of data: the Montreal travel survey, the national census, transit agency smart cards, and operational data including bus schedules and traffic signal timings. Each database is described in terms of atomic units of analysis, transportation objects and their attributes, as well as performance indicators resulting from the synthesis process. The analysis is performed using three types of technology: GIS, data visualisation methods, and an interactive tool for examining the raw data.

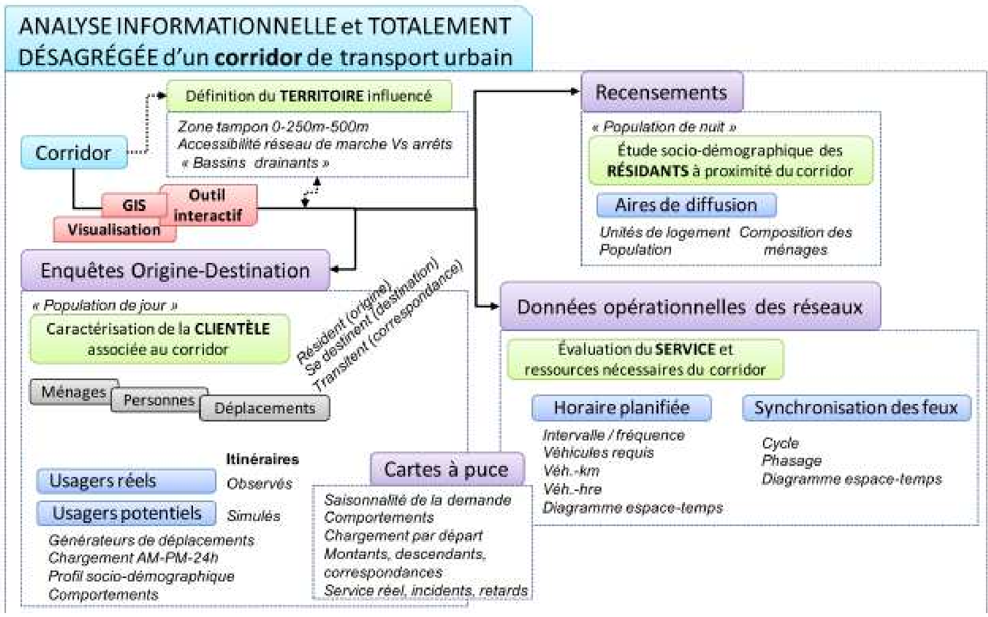


Figure 1.9: Application of the totally disaggregate information-based approach to the analysis of a transportation corridor in Montreal (Chapleau & Piché, 2009)

The salient feature of the totally disaggregate approach is that it is not based on a system of zones. Objects which exist in space (households, transit stops, trip origins, trip destinations) are all geocoded as precisely as possible. This approach obviates the need for centroids and associated connectors, as well as origin-destination matrices. An information system is constructed which manages and relates databases describing each of the objects simulated in a transport model. The simulation process itself is a heuristic for associating different databases with each other. For example, a traffic assignment model associates the database of road network links with the database of auto-drive trips.

1.3.3.1 The Montreal Travel Survey

The Montreal travel survey is a revealed preference survey based on telephone interviews of households. The first one was undertaken in 1970 by the Montreal Transit Corporation (*Commission de transport de la communauté urbaine de Montréal* or CTCUM) for the purposes of planning an expanded public transit network at a time when the public transit mode share was near 50% and when computing power was still very expensive. The survey consisted of interviews of roughly 5% of the households located in the region served by the CTCUM. As a consequence, the survey area was limited to a large section of the island of Montreal. A major network redesign, including a significant expansion of the subway system, was being planned to coincide with the 1976 summer Olympics, which Montreal would host. At the time, regional-level transit assignment models had not been implemented and so large quantities of data were sought to determine public transport usage patterns. The crucial question in the survey was "which line(s) did you take?" Related planning constructs, such as zone systems and networks, were designed *based on path information declared by the survey respondents* and the resulting data structure permitted a coherent and credible assignment of trips to the transit network (Chapleau, 1974). Subsequently, exponential increases in computer processing speeds and data storage capabilities had several important consequences. First of all, they have made possible the codification of the entire public transport network down to the level of individual bus stops. Also, public transit planning methods are no longer based on zones. Origins and destinations are now geocoded to the nearest metre. Aggregations, when necessary, are done based on major trip generators. This framework is sufficiently credible to be used as the basis for the distribution of fare revenue and provincial subsidies to each transit agency operating in the region.

Gradually over the past few decades, the survey has fallen under the shared jurisdiction of the *Société de transport de Montréal* (a direct descendant of the CTCUM), other regional transit agencies, and the provincial ministry of transport. The data it generates are no longer used exclusively for transit planning, but for a wide variety of analyses and research concerning all modes of travel. The survey area has been continually expanded and now constitutes a 4% to 5% sample of the entire island of Montreal, its close mainland suburbs and many outlying, almost rural, regions as well. In addition to asking transit riders which lines they used during their trip, the survey also asks drivers which freeways and bridges they used. As is already the case with transit line data, this information would ideally be used to estimate and validate traffic assignment models.

The Montreal travel survey has been refined over several decades and provides an extremely rich assortment of information on households, individuals and trips. When the survey data are combined with other available information such as census data and geomatic descriptions of transport infrastructure, statistical and spatial analysis techniques can be applied to address a very wide array of questions relevant to the urban transport planning discipline (Morency, 2004).

The survey is not perfect, naturally. Only one interview is performed for each household. The respondent answers on behalf of himself and all other household members, of whose travel behaviour he has only limited knowledge. The sample of interviewed households is likely biased in several respects, even though some controls are in place to ensure that each segment of the population is well-represented. The quality of the resulting data depends to a significant extent on the work performed by the interviewers themselves, none of whom are transportation specialists. Much effort has been devoted to the development of a graphical user interface which assists the interviewer over the course of the interview by maximizing the coherence of the input information (Chapleau, 1997).

1.3.3.2 Application to public transit systems

The development of the totally disaggregate paradigm in Montreal was initially motivated by the need to adequately simulate the public transit system. The travel behaviour of transit users is strongly influenced by the access and egress trip components (usually made on foot) and these could not be represented using the zone system required by the four-stage method. The goal has always been to perform detailed and multi-dimensional analyses of the transit system and its clientele. Complete descriptions of observed travel behaviour facilitate the planning of public transit services and the associated allocation of resources. These methods are currently used to plan bus routes and service frequencies, to determine optimal fleet sizes and to assess the equitable partitioning of the public transit subsidy across the various geopolitical entities whose residents benefit from the service to varying degrees. Totally disaggregate trip assignment is also used to validate the declared route information collected by the travel survey.

1.3.3.3 Application to public roads

The application of the totally disaggregate approach to the Montreal road network was investigated by Bergeron (1991). Analysis of the general interaction of travel demand and supply has required the development and maintenance of a detailed geomatic road network covering the entire metropolitan area. This network has been used in all-or-nothing assignments of automobile trips contained in the travel survey without recourse to a system of zones or an origin-destination matrix. Traffic congestion, however, is not explicitly represented. The absence of congestion is usually acceptable in public transit simulations since an increase in transit demand does not increase the cost of travel, as a general rule. Variations in road network speeds have nonetheless been modelled based on direct observation with GPS-equipped para-transit vehicles (Allard & Grondines, 2007). Since almost all existing models of road traffic congestion are based on the assignment of demand between zone centroids, a new method must be found for representing congestion within the totally disaggregate information-based approach. The development of such a method is one of the goals of this research.

1.4 Equitable Road Transport

The term public transportation is usually taken to refer to mass transit services provided by municipal governments. It should be pointed out, however, that almost all urban transportation, regardless of mode, is public transportation in the sense that it involves the act of sharing. Fundamentally, travelers in a bus or in their cars are sharing space. They are also sharing economic and natural resources. Any situation which requires sharing necessarily raises questions about fairness. In transportation studies, the word “fairness” is often represented by the term “equity” (Forkenbrock & Sheeley, 2004; Levinson, 2002).

Equity in urban transportation has many dimensions. In the ecological context, it refers to equity between generations: the current generation must preserve resources for the generations to come. These issues often fall under the headings “sustainable development” or “environmental justice.” There is also a notion of social equity arising from the realisation that not all socio-economic groups enjoy the same access to opportunity. Transportation in general and public transportation in particular is sometimes advocated as a measure to address the resulting inequalities in income and quality of life. Yet another perspective on equity invokes the user-pay principle whereby individuals who use a particular transport service should pay the full cost of their consumption. Within the domain of regional traffic models, the vast majority of research in this area has focused on this latter objective. It has gained additional traction because of its compatibility with the objective of intergenerational equity since both approaches aim to reduce automobile use (or at least make it more efficient).

Less often, the issue of equity is considered as a geopolitical problem. In most jurisdictions, transport infrastructure is funded through taxation by local, regional or federal governments. A government’s authority is defined by its administrative limits which are nothing other than boundaries in space (zones). The interaction between a network (transport infrastructure) and a system of zones (administrative regions) can result in the inequitable redistribution of resources as well as the evasion of fiscal responsibility by some jurisdictions. All these discussions are fundamentally concerned with a perceived distribution of costs and benefits among different groups. Ideally, the distribution would be “fair”, although the definition of fairness is highly subjective. Some degree of clarity can only be achieved if the costs and benefits in question can be quantified and if certain objectives of a transportation system can be agreed upon. In the context of the present research, the transportation system in question is an urban network which serves private automobiles. The next section of this chapter deals with issues of costs, benefits and purpose of such a system. The second section discusses the user-pay principle more commonly referred to as marginal-cost pricing. In the third section, equitable transport as a geopolitical problem is explored. The fourth section details the redistributive effects of a public transport system.

1.4.1 Costs, benefits and objectives of urban road transport

The costs of urban road transport are well-recognized. Indeed, the private car is considered a dangerous nuisance by many. The benefits of the road network are also obvious, although they tend to be de-emphasized in the current political climate. The role of the road transport system in the functioning of urban areas is also, to some degree, taken for granted.

Lakshmanan et al. (2001) note that the primary benefit of a transport system is derived not from its provision but from its use. This statement is based on the assumption that the demand for transportation is derived from demand for other economic opportunities. The authors also distinguish between “external” and “internal” costs and benefits. “External” refers to costs or benefits that are not included in the transaction price of the transport service while “internal” refers to costs or benefits that are included in the transaction price. Also, because of the interaction induced between otherwise insulated populations, the provision of transportation services often produces clear winners and losers:

The incidence of gains and losses over different interest groups may thus vary over space. This means that equity considerations, and issues of social feasibility are likely to be important determinants for the viability of infrastructure policies.

Lakshmanan et al. list the numerous benefits generated by a transportation system. These include economic stimulation during the construction of new infrastructure, the development of trade between regions; increased access to opportunities distributed over space and time, improved productivity, greater market competition and decreased collusion. The authors note that, with the exception of effects related to construction, these benefits should not be considered external since they all accrue to users of the system. External benefits of usage are considered to be negligible.

The costs of transport are also classified in terms of provision, usage, and whether they are external or internal. The provision of transport services generates construction and maintenance costs. These costs are usually internalized by government through taxation of the general population. The internal costs of usage include vehicle purchasing, maintenance, fuelling and insurance. There is also a cost associated with each trip but this cost is difficult to quantify monetarily since it is most easily calculated in terms of time which, although valuable, has no price. External costs of usage include traffic congestion, pollution, noise, annoyance and accidents.

The costs of traffic congestion provide an especially rich topic for discussion. Congestion is considered an external cost because individual drivers on a congested road do not pay for the delay they impose on other drivers. This premise forms a virtually universal consensus among economists. There is a related theory which goes further by stating that traffic congestion represents a loss to the economy in general (Arnott & Small, 1994; Conseillers ADEC Inc., 2009; Transport Canada, 2006; Weisbrod, Vary, & Treyz, 2001). This idea is based upon the assumption that the road transport system is not optimally priced and the inefficiencies which result represent a cost to individuals and businesses. Monetary estimates of this economic cost of congestion are founded on presumed values of time and are frequently used as justification for the widespread implementation of marginal cost pricing.

The notion that urban traffic congestion imposes a monetary cost on society raises important points about the purpose or objective of an urban road system. As suggested in a paper by Stopher (2004a), if the purpose of road infrastructure projects is to reduce congestion, then they are spectacular failures since traffic congestion everywhere has increased in lock-step with the quantity of new facilities. Fortunately, the purpose of most transportation projects is not to decrease traffic levels but to increase mobility. New infrastructure allows people to travel faster and therefore further within a fixed time budget (Zahavi, 1979). In this context, traffic congestion can be thought of as an indicator of the price people are willing to pay for access to a particular location. Efforts to reduce congestion are really efforts to reduce consumption, not just of transportation, but of all goods and services which are accessible by car. While the goal of reduced consumption may have merit in specific situations, the wealth-destroying measures required to maintain free-flow speeds on all roads at all times are never described in such terms by their advocates.

1.4.2 Marginal cost pricing

Historically, much of the discussion around the pricing of roads has centred on the difference between the marginal cost and the average cost of consumption. The average cost is a somewhat abstract notion describing a total cost that is distributed evenly among all consumers. The marginal cost is what each individual consumer would pay if the price were to cover the cost of production. In strict mathematical terms, the marginal cost function is the first derivative of the total cost function. To give a very general example, people are said to pay the average cost of their consumption when a service is provided free of charge by the government. Of course, "free of charge" does not mean that the good or service cost nothing to produce. It just means that the cost of production was covered by tax revenues which are, in principle, completely independent from the amount of service consumed. By contrast, consumers pay the marginal cost when they pay only for what they consume.

The accepted economic wisdom holds that average cost pricing leads to serious inefficiencies and distortions because consumers do not pay the true cost of their consumption. It has long been argued that traffic congestion is caused by the inefficient use of road capacity which occurs when road users do not pay the marginal cost of their travel patterns (Vickrey, 1969). The theory is that drivers pay the average cost of their travel patterns through the amount of time they spend travelling. At a particular instant, this cost is the same for all users of a given facility. Moreover, a vehicle arriving on a congested facility has a much greater effect on the travel time of other road users than a vehicle arriving on an empty or almost empty facility. In other words, if a continuous monotonically increasing supply curve is assumed, the marginal cost of travel is always higher than the average cost. The difference between the marginal cost function and the average cost function at a particular level of demand is the optimal toll. According to the prevailing economic wisdom, the toll raises the cost of travel from the average cost to the marginal cost and serves to reduce demand based on the willingness-to-pay principle which is represented by the downward-sloping demand curve. The justification for this type of marginal cost pricing (also known as congestion pricing) is that non-toll public road networks are economically inefficient. A typical toll-free highway, for example, carries a volume of traffic well below its capacity for most of the day. During peak periods, however, the highway may experience demand that exceeds its capacity resulting in traffic congestion.

To rectify the situation, many papers have been written advocating the application of dynamic tolls to road networks. Since demand varies significantly over the course of the day, the marginal cost of using a facility will vary accordingly. Proposed solutions include time-variable tolls (Arnott, de Palma, & Lindsey, 1993) or cordon-based tolls in order to reduce traffic congestion at locations and periods when congestion is particularly severe. A system of this type has been implemented in Stockholm. The simulation of time-varying tolls requires a dynamic model of traffic of the type which has yet to be successfully implemented at the scale of a large urban area. On the other hand, tolls which vary only by facility or geography lend themselves to analysis using the traditional traffic assignment models based on the four-stage approach. Such analyses are made possible by placing a monetary value on travel time, and replacing the standard link cost function with a function of generalized cost. The generalized cost is the sum of the toll and the volume-delay function evaluated at current traffic levels converted into monetary units via the value of time coefficient. An elastic demand function is often incorporated in the simulation. Mohring (2006) offers a good example of such an exercise.

An extensive body of work exists describing the design and impacts of different tolling mechanisms such as tolls based on distance, tolls based on time spent on the facility, congestion pricing, area-based tolls and cordon tolls (Dial, 2000; May & Milne, 2000; Vrtic, Schuessler, Erath, & Axhausen, 2007; Yang & Zhang, 2002). Distance-based and cordon tolls are easily simulated in a traffic assignment model since they require only that an additional cost be added to the links that form the tolled facility or the cordon. More complex tolling systems require modified algorithms. For example, an area-based toll charges drivers for a permit to enter a designated sector. The permit allows re-entry during a specified period (typically one day). The London congestion charge operates in this way. In order to model such a system, it is necessary to track vehicle trip chains. A modified traffic assignment algorithm which does this was proposed by Maruyama and Sumalee (2007). In the Montreal region, user equilibrium traffic assignment methods have been used to study a distance-based toll on a new bridge between the Island of Montreal and the suburb of Laval (PB Consult. Inc., 2002).

Road pricing systems exist in many jurisdictions where their application is limited to interurban freeways. Implementation of congestion-charging mechanisms in urban areas is comparatively rare, in part due to technological constraints but also because such measures have always met with a great deal of popular resistance. This resistance stems from drivers' belief that they already pay sufficiently, through taxes, licensing fees and insurance premiums, for the privilege of driving. Moreover, non-tolled alternative routes or cheaper modes of transport are not available to many travellers. In addition, the calculation of an optimal toll requires the conversion of travel time into units of money, which is a far from self-evident exercise when applied to a large and diverse population (Atkins, 1984). Crozet and Marlot (2001) have even argued that the tolling mechanisms usually proposed will in fact do little to reduce congestion in urban areas since the price elasticity of auto-travel demand is extremely low. These issues have contributed to the discussion of equity in transportation because, under a conventional tolling mechanism, people with lower incomes are effectively priced off the road. The possible side-effects of this reduced mobility on already vulnerable populations have been discussed at length (Bonsall & Kelly, 2005; Lari & Iacono, 2006; Litman, 2006; Trannoy, 2006). As a result, tolling strategies are often classified as regressive and inequitable.

The possibility of reducing congestion on a particular facility through the imposition of a toll raises important questions about how the costs and benefits of a particular tolling scheme should be calculated. If congestion is reduced on the newly-tolled road, it is because some people choose to alter their behaviour rather than pay the toll. This change could take several forms: if the toll varies with time, a driver could choose to travel at a time when he finds the price more reasonable; the driver could switch routes and avoid the toll altogether; the driver could switch modes; the driver could choose not to travel at all. In order to analyse the possible impact of a toll regime, it is necessary to identify these distinct markets. The identification process necessarily entails the collection of information characterizing drivers and their trip-making behaviour. Relevant attributes include whether or not trips are discretionary, scheduling constraints, traveller income and the accessibility of alternative routes and modes. These details are often ignored in the economic evaluation of road pricing mechanisms.

1.4.3 Equitable Transport as a Geopolitical Problem

In large metropolitan areas, publicly financed road and transit networks both serve multiple jurisdictions, each of which is expected to pay a share of the required subsidy. In order to minimize problems of fiscal evasion and economic distortion, it is necessary to evaluate, for each geopolitical entity (city, municipality and borough), the costs and benefits it incurs relative to its financial contribution to the transport system. Analyses of this type are rare in the literature. In the Greater Montreal Area, travel survey data have been used to measure the road transport consumption patterns of multiple population groups differentiated by socio-demographics and geography (Essakali, 1999). The observed variation of consumption patterns suggests an infrastructure financing mechanism based on population attributes. The most practical approach exploits the differences in travel consumption between geopolitical entities. A methodological framework of this sort, based on the detailed information contained in the Montreal travel survey, has been adopted to structure cost sharing agreements among the many municipalities whose residents benefit from intra-regional public transit. Costs per jurisdiction are measured in terms of passenger-km supplied and benefits are represented by passenger-km travelled. An example of the methodology is found in (Chapleau, 1995).

1.4.4 Redistributive effects of a transport network

A defining characteristic of a transportation network is the way in which it redistributes costs and benefits over space and time. A prime example is the phenomenon of urban sprawl and the hollowing-out of city centres. This phenomenon is driven, at least in part, by the ability of individuals to avoid paying the high costs of life in the central city by moving to the low-cost suburbs. In the suburbs, the benefits of the urban agglomeration are still accessible thanks to the transport network whose costs are borne primarily by the central city. The quantification of such effects relies upon an accurate identification of infrastructure *users*, something which is not easily accomplished using the traditional four-stage model. Almost all traffic assignment platforms offer the possibility of performing a “select link analysis” which outputs the set of links and link flows which feed a link or set of links chosen by the user. The set of o-d pairs which generate the demand for the selected links is also output. However, important information describing the demand – particularly trip purpose, departure time and personal and household attributes – is lost.

In the Montreal context, the application of the totally disaggregate approach permits an analysis of infrastructure clientele along multiple dimensions. Chapleau and Morency (2004, 2005) examined the consumption of space, time and transportation resources in an effort to reveal the economic distortions arising from various transport policies and evolving societal trends such as the ageing population and a growing level of auto-ownership. The study reveals that major transportation facilities (such as the metro system and the bridges linking the island to the city's suburbs) have distinct usage profiles, particularly from a geopolitical standpoint. For example, the bridges are used much more frequently by off-island residents than by Montrealers.

Air pollution is among the important costs redistributed by the road network. Traffic volumes and speeds are correlated with concentrations of atmospheric pollutants. Efforts to model this correlation on the Island of Montreal have been made using the results of a static traffic model (Crouse, Goldberg, & Ross, 2009). The resulting estimations of nitrogen dioxide concentrations are shown in Figure 1.10. The traffic model used the Montreal travel survey to represent demand and a complete digitized road network to represent supply (Spurr, 2005). A subsequent analysis compared observed levels of air pollution with indicators of deprivation at the neighbourhood level and revealed a complex relationship between the two (Crouse, Ross, & Goldberg, 2009).

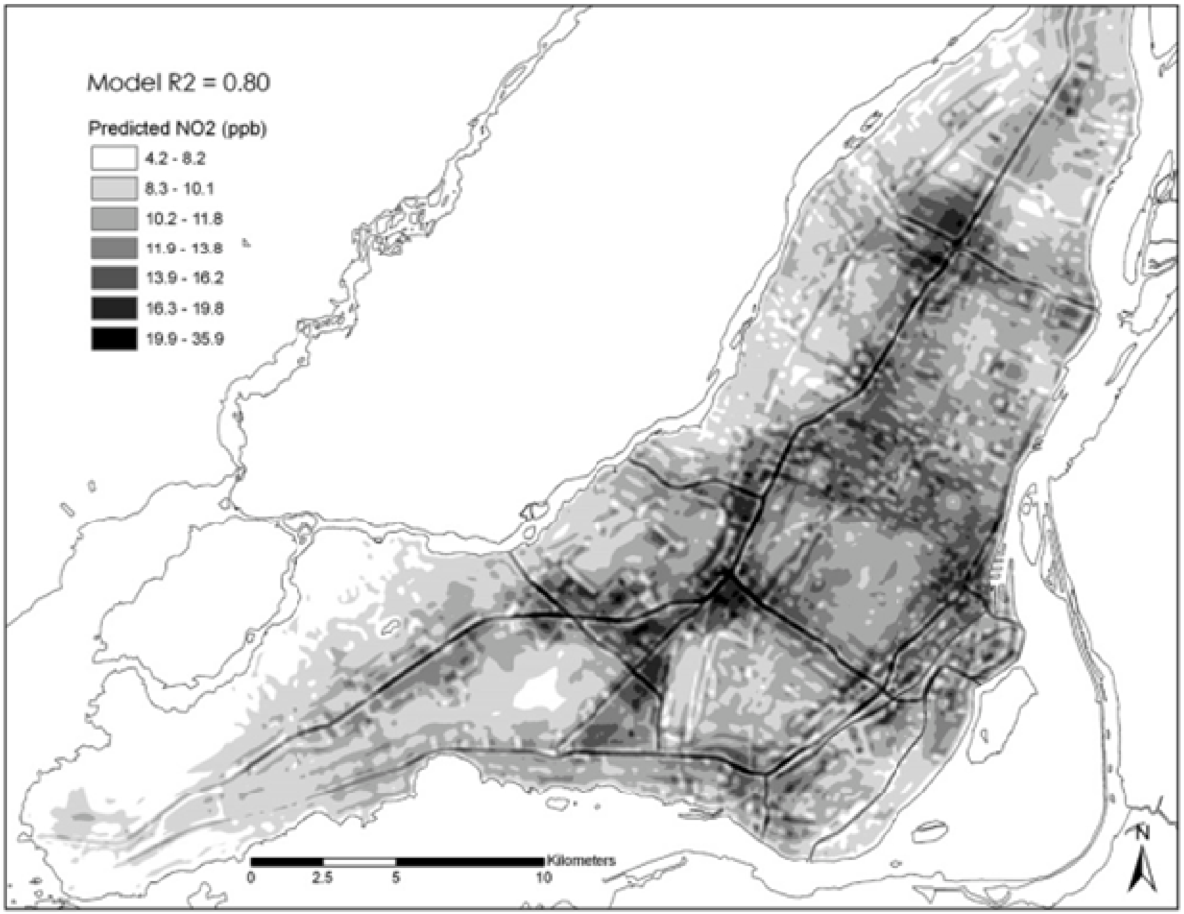


Figure 1.10: Nitrous oxide concentrations predicted using a static traffic model (Crouse, Ross et al., 2009)

1.5 Conclusions

The tools and methods employed in the planning of public road systems have evolved considerably over the last few decades. Highly simplified representations of urban road systems are now accompanied by detailed and complex models of individual vehicles and people. While the complexity of human behaviour may justify elaborate artificial constructs, the algebraic intricacy of these platforms is a barrier to transparent (and sometimes honest) discussion especially once the models are put to use outside the academic realm. Increasing complexity, however, is not the primary challenge faced by transportation planners. In fact, the creation of new formulae and algorithms is simple compared to the task of deriving useful information from the enormous quantities of data which are increasingly available. An examination of the academic literature reveals that this challenge has yet to become a priority for researchers even though private interests are already capitalizing on the power derived from information management.

The traditional transport planning paradigm has been used to analyse questions of equity in all its forms. Travel behaviour is understood to be the result of an economic equilibrium which is policy-sensitive and much work has been devoted to the development of models which incorporate various pricing mechanisms. Despite these efforts, lack of information and dependence on complicated yet simplistic algebraic methods have contributed to the absence of effective initiatives for reducing traffic congestion, limiting urban sprawl and curbing the consumption of non-renewable energy.

CHAPTER 2 DEVELOPMENT OF A DISAGGREGATE INFORMATION-BASED ROAD NETWORK MODEL

By convention, urban traffic models are designed to predict vehicle volumes and congestion levels on network links. In addition, these models are frequently employed for the estimation of travel times during periods of peak demand. The goal, in principle, is to test the effect of a particular transportation policy on the transportation system as a whole. Average traffic speeds and flow rates are the most commonly adopted consumption metrics. From a predictive modelling perspective, however, it can be argued that these indicators are of secondary interest since any changes in speed or flow are an aggregate result of a change in travel patterns which may not be represented explicitly in the model. In other words, the performance of an element of road infrastructure is less important than a proper identification of its actual and potential users. The statement is especially true when evaluating policies relating to equitable pricing and financing of transport services.

In the past, technological constraints placed severe limits on the ability of traffic modellers to construct anything other than greatly simplified models of urban road systems. Now that these constraints have largely disappeared, there is a consensus among the modelling community that more sophisticated methods should be developed. Many of the proposed new methods were described in the previous chapter and, while interesting and highly relevant, in many cases their complexity makes them hard to operate and validate, particularly in situations where data are scarce. Moreover, the construction of mathematical algorithms often takes precedence over the analysis of real information. It is difficult to evaluate the applicability of any modelling method in such a context. Nevertheless, many of the “avant-garde” models incorporate two worthy principles. The first principle is the notion of a microscopic model where travel is analysed at the level of individual persons. Such an approach is justified on the grounds that travel is a human activity and that each human exists in a unique environment. The second worthy principle is the idea that a transport model should be dynamic. Within a single day, the urban transportation system evolves considerably over time and its time-varying characteristics should be explicitly represented since they constitute an important feature of the traveller’s decision-making environment.

The intention of this research project is to develop a model of an urban road network which is microscopic and dynamic. Because of the institutional context in which this research was carried out, it is convenient and appropriate to note that the microscopic-dynamic paradigm can be encapsulated within the totally disaggregate approach to transportation planning. Although aggregation is usually required to present meaningful results, the analysis invariably begins with a treatment of the observed behaviour of individuals at precise moments in time and locations in space.

In the experiments described below, the observed individual behaviour consists of a sample of the Montreal travel survey for which partial information on the choice of route has been collected. This partial information is composed of precisely geocoded origins and destinations, as well as a major road facility (bridge) used to complete the trip. The goal at the outset is to find modelling method which can reliably reproduce the observed facility choices. The more important objective, however, is to present an experimental methodology for identifying and characterizing the users of major road infrastructure based on the treatment of detailed information. Such a method is a prerequisite for the quantitative analysis of equity in road transport.

The proposed methodology – dubbed information-based disaggregate traffic assignment – is summarized in Figure 2.1. The left side of the figure contains all the objects which can be fully described using readily available data. The study region is defined as the Greater Montreal Area which is composed of numerous geopolitical entities (cities and boroughs). These entities exist to the extent that they are populated and partially financed by tax-paying households composed of individuals. Most of these individuals travel and an important majority do so by car. Each auto trip is characterized by an origin, a destination and a sequence of road infrastructure facilities used to complete the trip. The road facilities which are the focus of this research are the fifteen bridges which provide access to the island city of Montreal. Together, all these elements represent the observed demand for automobile travel. The supply of transport consists of all the road facilities in the region, classified by function and jurisdiction. Note that the characteristics of supply and demand are time-dependent (dynamic), because of the distribution of traveller departure times and the evolution of traffic conditions.

The right side of the figure describes the artificial constructs which are the product of a model. For simulation purposes, an artificial network is constructed using the concepts of links, nodes and movements. The simplified representation of supply is associated with the observed demand using a hypothesis about the route-finding behaviour of drivers. This hypothesis forms the core of the traffic assignment/route choice simulation. It generates as output itineraries of individual trips. These simulated paths can be subsequently aggregated in order to calculate traditional measures of consumption such as link flow and average speed. These metrics contribute to the estimation of route choice parameters if an iterative feedback mechanism is adopted. They also allow for the quantification and distribution of the costs and benefits associated with the current travel patterns. The assignment of these costs and benefits to the geopolitical entities of the region as well as their constituent households constitutes the analysis of equity which is discussed in the next chapter.

This chapter proceeds as follows. Section 2.1 presents a brief discussion on the tools used to complete the research. Section 2.2 describes the study area and provides a brief description of the 15 bridges which are the focus of the research. Section 2.3.2 discusses the representation of the regional road transport system. Section 2.4 describes the sample of survey data describing the usage of these bridges. Section 2.5 describes the development of two totally disaggregate models of facility choice. Section 2.6 summarizes the findings of the experiment.

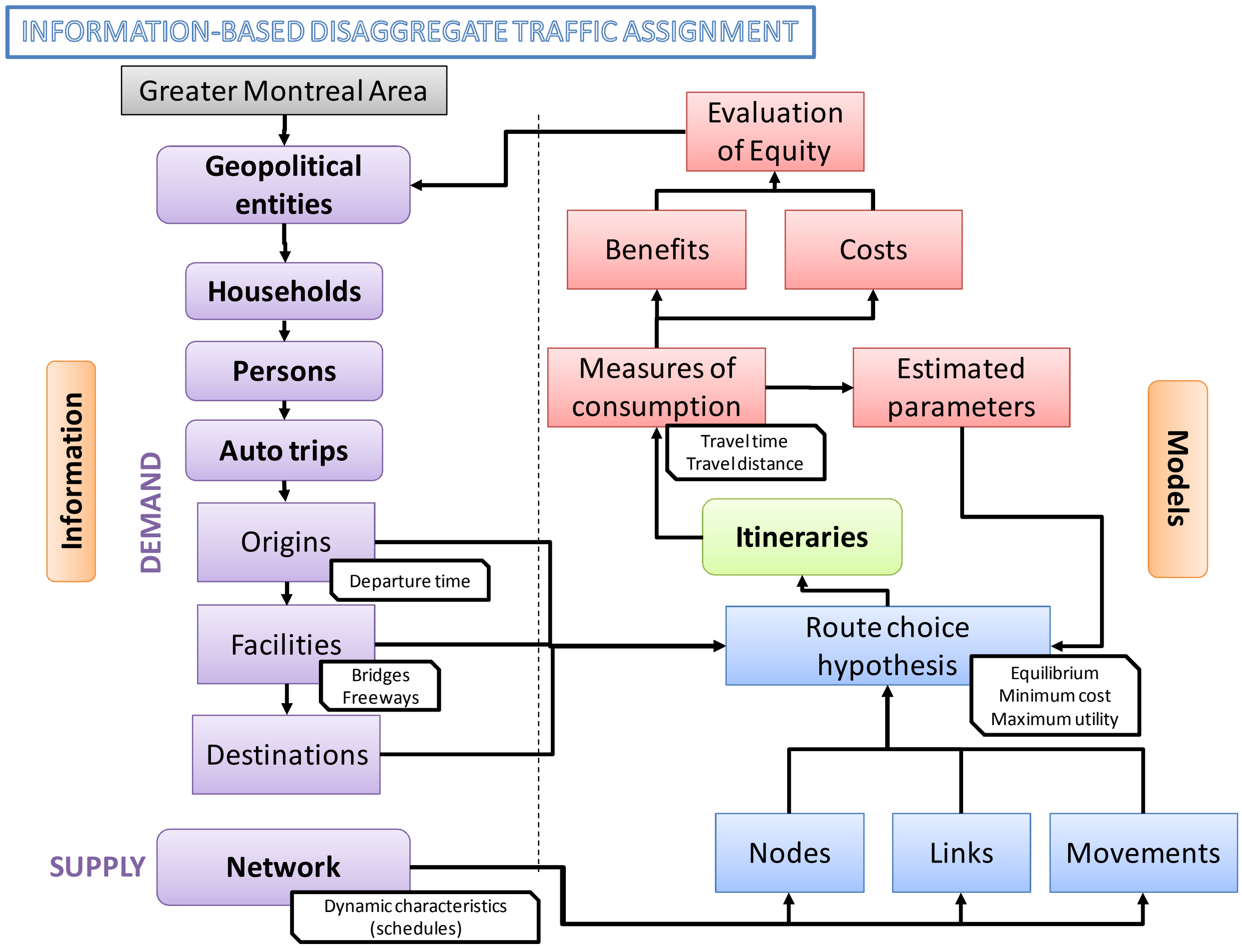


Figure 2.1: Schematic of the proposed model structure

2.1 “Open-source” Transportation Analysis Tools

If transportation planning is to be considered a true science, then an experimental methodology must be devised for the application of the scientific method. Field experiments are impractical or impossible because of the formidable quantities of people, equipment and money they would require. Experimentation in a virtual environment is the only feasible option. The construction of a virtual environment in which scientific exploration (or practical application) can be performed is not necessarily straightforward. It requires the proper installation of software tools and the efficient management of input and output. It also requires of the modeller a fairly complete comprehension of the mechanisms by which the tools he is using function. Commercial software is usually designed in a way which hides its inner workings. This is done in part to make the program "user-friendly" and also to protect proprietary information. Over the last several years however, an exponential reduction in computing costs combined with an explosion of information transmitting capacity has greatly augmented the level of analysis that can be performed for very little monetary expense. Many computer programs are now "open source" meaning that their source code can be obtained and modified free of charge (for details, see http://www.opensource.org/). These tools, while occasionally more limited and difficult to use than their commercial counterparts, constitute a very effective method for the preservation and transmission of knowledge. Although they are designed with practical considerations in mind, their pedagogical potential makes them especially valuable in an academic context. Moreover, their combined use often yields an experimental environment which is more productive than one provided by single commercial software.

Of particular interest to urban transportation planners are geographic information systems (GIS), statistical software, database management, transportation simulation packages and transportation network data. An array of such tools was used in the present research and they, as well as some others not directly employed, are presented in Table 2.1 below.

Table 2.1: Open-source transportation analysis tools

Type	Software	URL
Geographic information systems	OpenJump fGIS DIVA SAGA PostGIS QGIS	www.openjump.org www.forestpal.com/fgis.html www.diva-gis.org www.saga-gis.org http://postgis.refractions.net www.qgis.org
Statistical analysis	R Biogeme	www.r-project.org http://biogeme.epfl.ch
Database management	PostgreSQL	www.postgresql.org
Transportation planning	GeoDA CrimeStat TRANSIMS MATSim CiudadSim	http://geodacenter.asu.edu www.nedlevine.com/nedlevine17.htm www.transims-opensource.net www.matsim.org http://www-roc.inria.fr/metalau/ciudadsim/
Transportation network data	GEOBASE OpenStreetMap	www.geobase.ca www.openstreetmap.org

2.2 Definition of the study area (Greater Montreal)

The Greater Montreal Area (Figure 2.2) is a metropolitan region covering approximately 10,000 square kilometres and inhabited by roughly 4 million people. The region is located at the confluence of two major waterways: the Ottawa River and the St. Lawrence River. The City of Montreal is an island just east of the point where these two rivers meet. Immediately to the north of Montreal is a smaller island which is fully occupied by the suburban city of Laval. The small river which separates Montreal from Laval is called the Rivière-des-Prairies. Across the St. Lawrence from Montreal are a number of municipalities known collectively as the South Shore. Beyond Laval and the South Shore are exurban regions known as *couronnes* (crowns). The *couronne nord* encompasses municipalities to the north of the Ottawa and St. Lawrence rivers. The *couronne sud* is everything to the south of the Ottawa and to the east of the St. Lawrence, except for the municipalities within the South Shore.

The region is served by a road network with a centre-line length of roughly 20,000 kilometres, 1,600 of which are classified as freeways. The road network of the city of Montreal is connected to the rest of the North American road network by 15 bridges. Ten of the bridges carry roads which are easily classified as freeways. Four bridges carry urban arterial roads. One bridge, the Jacques-Cartier, is difficult to classify.

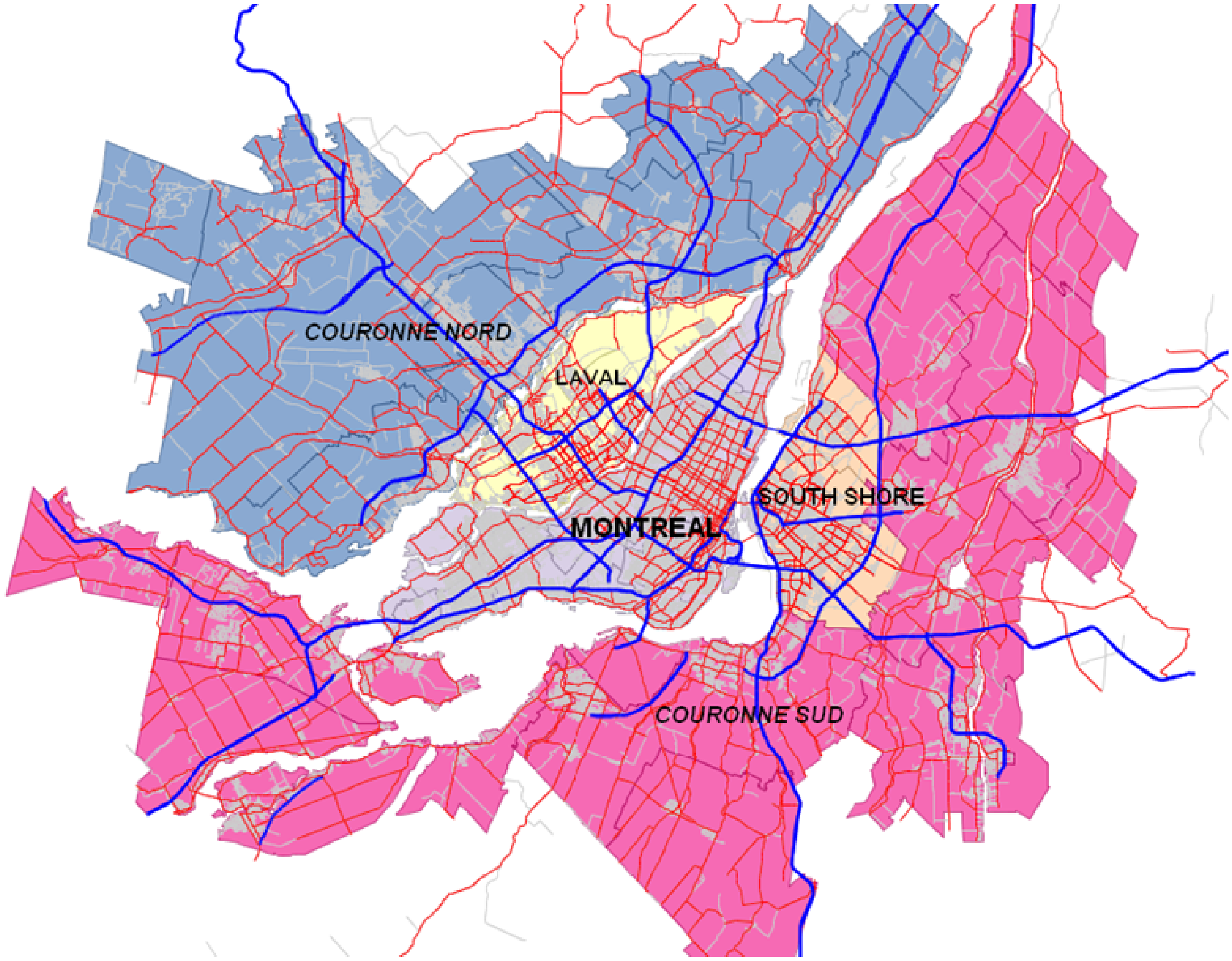


Figure 2.2: The Greater Montreal Area and road network

2.3 The regional road network and its representation

The construction of a simulation model of regional traffic patterns requires a correct faithful representation of transport supply embodied by the road network. The quality, quantity and structure of the data used to construct this representation are all important aspects of the modelling process. The present study is primarily concerned with the traffic on the major bridges and so these facilities are described in some detail in the first sub-section. The second sub-section describes the data on which the model network is based. The third sub-section discusses the numerous elements which are considered in the codification of the road network of a large urban area.

2.3.1 The Major Bridges of Montreal

The geographic layout of the 15 bridges providing access to Montreal is represented by the simplified graphic shown in Figure 2.3. Each bridge belongs to one of 4 screenlines defined by the network of rivers and lakes along the axis of the St. Lawrence River. The four screenlines together constitute a cordon around the city of Montreal. When trips using multiple bridges are excluded, each screenline becomes largely independent of the others. For example, the Laval screenline will not be crossed by anyone seeking to access Montreal from the South Shore. There are some trips, however, where a choice between two screenlines is possible. The two bridges forming the "East" screenline are occasionally used by travellers originating in the *couronne nord*, for whom the bridges of the Laval screenline are also feasible options. Moreover, the Mercier Bridge (1501) and the two bridges of the West screenline are all possible alternatives for a handful of trips originating to the southwest of Montreal. These screenlines and their composite bridges are examined in some detail.

2.3.1.1 The Laval Screenline

The Laval screenline runs along the Rivière-des-Prairies which separates the island of Montreal from the island of Laval. It is composed of six bridges. All the bridges of the Laval screenline carry roads controlled by the provincial (Québec) government.

Pont Viau (1401): This 1 km-long bridge carries provincial highway 335 but the functional class of the road is an urban arterial with two lanes in each direction. Access to the bridge at both ends is controlled by traffic signals.

Pont Papineau (1402): The Montreal end of this bridge is a terminal point of Autoroute 19 which is a short freeway running north-south through Laval. The bridge carries three lanes in each direction over a distance of 1.2 km. A traffic signal controls access at the Montreal end.

Pont Pie-IX (1403): This bridge, roughly 1 km in length, is the easternmost facility in the Laval screenline and, like the Pont Papineau, it carries a six-lane freeway (Autoroute 25) which terminates at the Montreal end. The freeway resumes a few kilometres to the east and runs to the Lafontaine Tunnel (1304).

The Pont Lachapelle (1404), 1.1 km long, is similar the Pont Viau. It carries a provincial highway (Rte 117) but the road is in fact an arterial with three lanes in each direction. Traffic signals control access at both ends of the bridge.

The Pont Médéric-Martin (1405) carries the very busy Laurentian Autoroute (A-15) which begins about 150 km north of Montreal and serves many suburban communities. At the point where it crosses the bridge, the freeway has four lanes per direction. The bridge spans 1.5 km.

The Pont Louis-Bisson (1406) is 1.3 km long and carries Autoroute 13 which runs parallel to the A-15 between Montreal and a point just north of Laval. At this location, the freeway has 4 lanes going north and 3 lanes going south.

2.3.1.2 The South Shore screenline

The South Shore screenline is defined by the St. Lawrence River and seaway which separates the Island of Montreal from the Québec mainland. Near Montreal, the width of this watercourse varies between a few hundred metres and several kilometres. Five crossings have been built at some of the narrower passages. By Canadian law, any road bridge which passes over the St. Lawrence Seaway shipping lane falls under the jurisdiction of the federal government (the government of Canada). The Champlain, Mercier and Jacques-Cartier bridges are all in this category. Although the Victoria Bridge is owned and operated by the Canadian National Railway, maintenance of the road portion of the bridge is reimbursed by the federal government[1]. The Lafontaine Tunnel runs underneath the seaway and is under the jurisdiction of the province of Québec.

The Pont Champlain (1301): According to the MTQ's freeway numbering system, this extremely busy bridge carries three freeways: the A-10, the A-15 and the A-20. Physically, it is a six lane freeway facility spanning 4.5 km with large interchanges at both ends. During peak periods, one

---

[1] Public Works and Government Services Canada authority code A411 (http://www.tpsgc-pwgsc.gc.ca/recgen/pceaf-gwcoa/0809/txt/rg-5-code-info-1-11-a-a2-a41-a411-eng.html):

This statutory authority (pursuant to Vote 107, Appropriation Act #5, 1963) provides for expenditures related to the Victoria Bridge for payments to CN Rail for loss of toll revenue and for rehabilitation work on the roadway portion of the bridge.

of the lanes is reserved for public transit buses which run against the flow of traffic. At such times, therefore, the bridge carries 2 lanes in the non-peak direction and 3 lanes in the peak direction.

The Pont Victoria (1302) is a 3.9 km-long railway bridge constructed in the 19th century. Two automobile lanes were added several decades after its completion. During regular operation, each lane serves a different direction. During peak periods, however, both lanes flow in the direction of prevailing traffic. This means that both lanes flow inbound to Montreal during the morning rush hour. The Victoria Bridge carries provincial route 112 which is best classified as an urban arterial. Traffic signals control access to the bridge at both ends.

The Pont Jacques-Cartier (1303) spans 3.1 km and is a peculiar species of bridge. It carries provincial Route 134 which is a signalized arterial road through most of its length within Greater Montreal. However, for a few kilometres on the South Shore leading to the bridge, it is a divided freeway with limited access and no traffic lights. The bridge is directly connected to another freeway (the A-20/132) at its South Shore end. The Montreal end is linked to the dense and congested arterial road network of the downtown core. On the bridge itself, Route 134 consists of five lanes with no physical separation between directions. The middle lane is reversible so, like the Champlain Bridge, the Jacques-Cartier has three lanes flowing in the peak direction and two in the non-peak direction. In addition to linking the South Shore with Montreal, it also provides access to and from an island which separates the seaway from the St. Lawrence River.

The easternmost facility in the South Shore screenline is the Hippolyte-Lafontaine Tunnel (1304) which carries Autoroute 25 over a distance of roughly 3 km. Because it passes underneath the seaway, it is the sole responsibility of the provincial government.

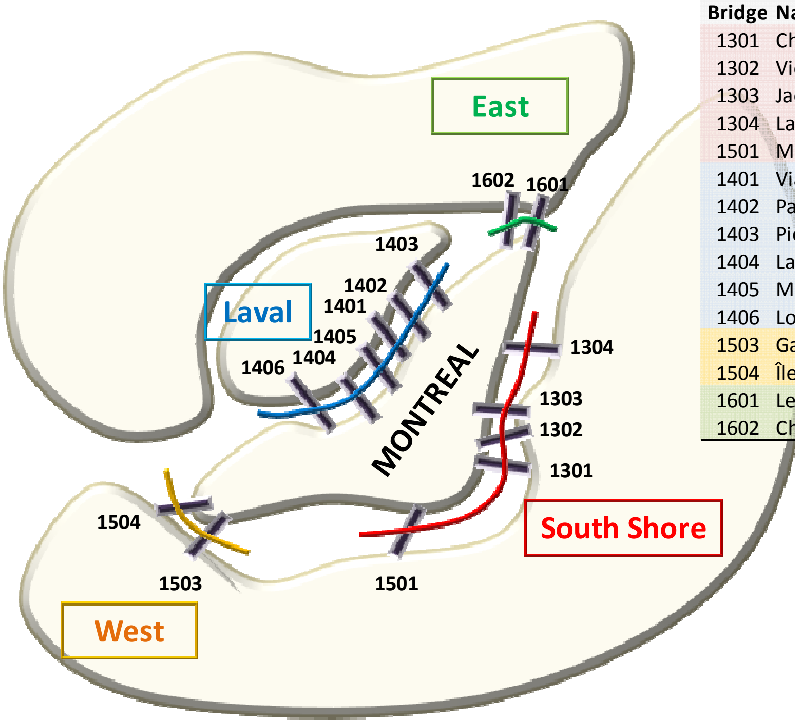
The Pont Mercier (1501) is also 3 km-long and carries two lanes of traffic in each direction. The road itself is provincial route 138 which is a freeway on the Montreal side of the bridge but becomes a multilane highway on the South Shore. The urban development at the South Shore end of the bridge is sparse meaning that the approach to the bridge is normally free of congestion effects, apart from those caused by the bridge itself.

2.3.1.3 The West Screenline

The Pont Île-aux-Tourtes (1504) is a six lane freeway facility carrying Autoroute 40 over a distance of 4.2 km. The freeway extends hundreds of kilometres from the bridge in either direction. The nearby Pont Galipeault (1503) also carries a six-lane freeway (Autoroute 20) but the road becomes a signalized urban boulevard just to the west of the bridge, which has a length of approximately 800 m. Both these bridges are the responsibility of the provincial government.

2.3.1.4 The East Screenline

The Pont Charles-de-Gaulle (1602) carries the six lanes of Autoroute 40 over a distance of 2.3 km. It runs parallel to the Pont Le Gardeur (1601) which carries a four-lane signalized urban arterial (provincial route 138) and spans almost 2 km. Both these bridges are the property of the provincial government.



Bridge	Name
1301	Champlain
1302	Victoria
1303	Jacques-Cartier
1304	Lafontaine
1501	Mercier
1401	Viau
1402	Papineau
1403	Pie-IX
1404	Lachapelle
1405	Médéric-Martin
1406	Louis-Bisson
1503	Galipeault
1504	Île-aux-Tourtes
1601	Le Gardeur
1602	Charles-de-Gaulle

Figure 2.3: The major bridges of Montreal and associated screenlines.

2.3.2 Network data

The fifteen major bridges constitute important components of the regional road network. The usage patterns of the bridges are influenced by characteristics of the larger network which must therefore be represented in a model of bridge choice.

One of the greatest challenges facing urban transportation modellers is the appropriate coding of networks. The coding of large urban road networks has traditionally been done by hand and, even in greatly simplified representations, it normally requires hundreds of person-hours to complete. The accuracy of the work is difficult to evaluate systematically. These problems have become even more acute with advent of sophisticated platforms which require an enormous quantity of detail such as traffic signal plans, coordination schemes and lane geometries in their representations of transport supply.

Ideally, networks would be coded automatically. Although such a procedure was impossible in the past, automated methods are becoming increasingly feasible due to the development of geographic perception tools. Satellite photos and in-vehicle GPS permit the rapid collection of large amounts of digitized information which describes the existing infrastructure. Numerous software platforms can convert geographic data (in the shapefile format, for example) into a logistic network which is usable for transportation simulation purposes.

While the incorporation of geometry can be made easier using automated methods, incorporating the connectivity of network elements is more complex. One-way streets and turn prohibitions must be included, respectively, as link and node attributes. Such information must either be inventoried by the institution responsible for the infrastructure or collected through on-site observation. Many of the corporations involved in mapping adopt the latter approach in the absence of governmental collaboration.

Digitized networks for the purposes of traffic assignment are most often simplified representations of real networks since the role of local roads is too minor to be of interest in a long-term forecasting exercise. It is important to note, however, that the local network represents the vast majority of lane kilometres in any urban area. Moreover, detailed representations of urban road networks are now easily obtainable. They also obviate the need for arbitrary artificial constructs such as centroid connectors. In keeping with Herbert Simon’s principle that it is the complexities in the environment that make human behaviour difficult to predict and since the necessary means exist, the detail in the network is included in the work presented below.

Two networks coded from independent sources were appropriated for this research project. The first was developed internally at École Polytechnique de Montréal. It is hereafter referred to as the Poly network. It contains 509 005 unidirectional links. This figure is much larger than the number of distinct road segments because a new link is drawn for each change in the road centreline trajectory. The second network was produced by the Canadian Ministry of Natural Resources and is available free of charge on-line at www.geobase.ca. This network is referred to as the GEOBASE network and is composed of 116 567 bidirectional links. Neither of these networks contains information on lane geometry or control systems. These elements are essential for the microsimulation of traffic.

Both the Poly and GEOBASE networks have a hierarchical structure (Table 2.2). Although comparisons between the two databases are easily made, a comparison with reality is more difficult since the levels of hierarchy in the real network are hard to distinguish using purely technical criteria. Despite completely independent codification methods and slightly different geographical coverage, the two databases are fairly consistent with each other. The largest differences are apparent in the length of the arterial and collector networks. The discrepancy is to be expected since a formal method for distinguishing between these two road classes remains elusive. Regardless, the combined arterial and collector networks are around 7000 km long in both networks.

Table 2.2: Hierarchical composition of two independent model road networks for Greater Montreal.

	Centreline kilometres	
Link class	**POLY**	**GEOBASE**
FREEWAY	1615	1240
RAMP	692	812
ARTERIAL	5023	1446
COLLECTOR	2079	4853
LOCAL	11639	10354
NON-AUTO		27
TOTAL	**21047**	**18732**

In addition to functional classes, the GEOBASE network contains route numbers assigned by the provincial government to roads within its jurisdiction. The route number attribute permits the identification of provincially-owned infrastructure. Using GIS tools, this information was transferred from the GEOBASE network to a simplified version of the Poly network using TRANSIMS module designed for that purpose. Link speeds under free-flow conditions were assumed to be equal to the posted speed limit. Speed limits were imputed according to link functional class and, in specific locations, by verification using Google Streetview. The end result of this processing is summarized in Table 2.3 describes the functional and jurisdictional hierarchies within the regional road network.

Table 2.3: Functional class composition of jurisdictional networks

Jurisdiction	**Functional class**	**Speed limit**	**Number of links**	**Length (directional km)**	**Network composition**
FEDERAL	FREEWAY	70	4	11	0.0%
	ARTERIAL	50	4	8	0.0%
			8	**19**	**0.1%**
PROVINCIAL	FREEWAY	70	198	131	0.3%
		80	13	9	0.0%
		100	998	1465	3.9%
	RAMP	50	1822	591	1.6%
	ARTERIAL	50	12	6	0.0%
		60	4702	2375	6.4%
		70	8	8	0.0%
	COLLECTOR	50	395	222	0.6%
			8148	**4805**	**12.8%**
MUNICIPAL	RAMP	50	721	172	0.5%
	ARTERIAL	60	12836	6920	18.5%
	COLLECTOR	50	8678	3820	10.2%
	LOCAL	40	71978	21664	57.9%
			94213	**32576**	**87.1%**
ALL			**102369**	**37400**	**100%**

2.3.3 Considerations in the codification of simulation networks

A civil engineer considers a road network from a physical perspective: as a collection structures, earthworks and other technological interventions that allow for the safe and organized movement of traffic. A driver, on the other hand, considers the road network as some artificial cognitive structure such as a sequence of instructions or a simplified map. In order to construct a meaningful model of automobile traffic, both physical and artificial conceptions must be considered.

2.3.3.1 An artificial hierarchical network for modelling driver behaviour

The available data make definitive statements about driver perceptions difficult. Since a detailed discussion of human cognition is beyond the scope of this research, the construction of an artificial road network is based on a simple but defendable hypothesis concerning the way drivers imagine the network. It can be safely asserted that very few drivers have complete knowledge of the network of a large city. A long-time resident, however, will certainly be familiar with most of the freeways, the principal tunnels and bridges, and many of the major roads. His knowledge of local roads is limited to parts of the network that he uses regularly. A hierarchical structure therefore suggests itself to the artificial network.

This approach is not new. It is commonly adopted in models of public transit networks because public transit systems have an explicit hierarchical structure whose levels are defined by the rapidity and/or capacity of the services offered. At the lowest level is a network of pedestrian facilities consisting of sidewalks, tunnels and walkways which provide access to the public transit service. Higher-level services come in many varieties ranging from the local bus to the express train and many variants in between. There is evidence to support the hypothesis that travellers choose paths which maximize the portion of the journey (measured in distance or time) spent on the most superior network possible. Moving from one level in the hierarchy to another usually involves a penalty measurable in time or money or both. Penalties are also incurred for transferring between lines within the same level (for a detailed discussion of hierarchy in transport networks, see van Nes, 2002).

In the present experiment, an analogous approach is adopted for auto travel (Figure 2.4). Drivers do not make long journeys using the local network but tend instead to seek out the superior network. They adopt this behaviour because travel on the superior network is more enjoyable and because their knowledge of the superior network tends to be more complete. In the conceptual model developed here, therefore, each automobile trip is subdivided into "access", "line" and "egress" components. The access and egress components are undertaken on the local road network while the line portion consists of the major roads and freeways which form the superior network. The analogy can be extended further to the application of Montreal travel survey data. Public transit users indicate which bus or metro lines they used to complete their trip and drivers are asked which bridge and freeway they used. In both the transit and road cases, the line information can be used to validate a simulation methodology (see section 2.4).

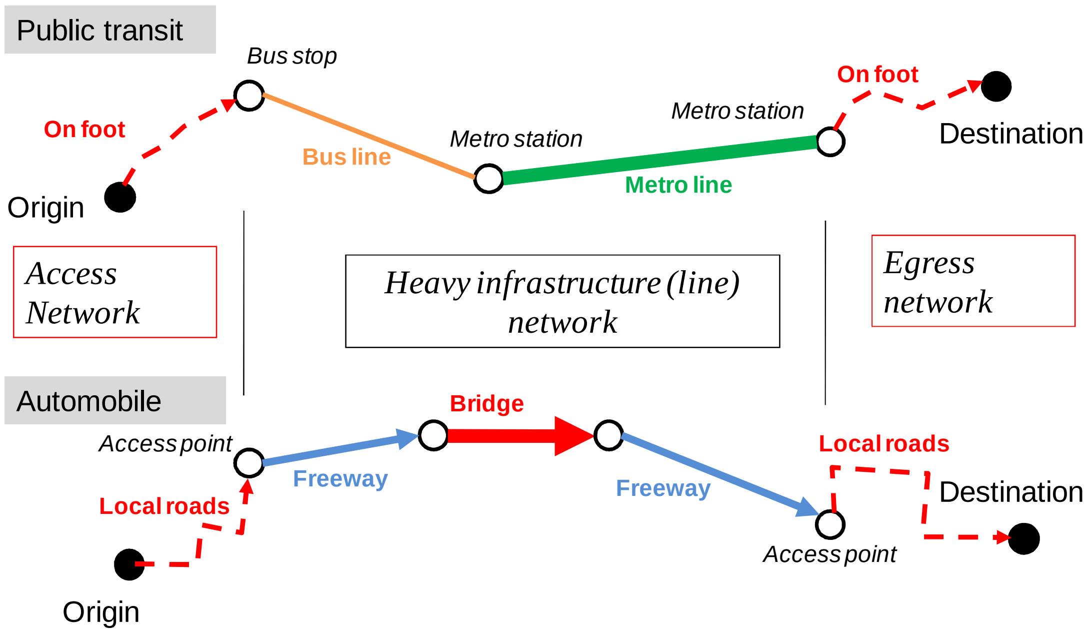


Figure 2.4: Conceptual route choice model based on the public transit network representation

This approach can be partially justified by the fact that a hierarchy does exist in the road network even if it is less rigidly defined than in the transit network. In Québec, as in other jurisdictions, the freeway system is subject to formal method of classification which is illustrated in Figure 2.5. All North American freeway systems use similar conventions. Each freeway is a bidirectional facility with defined start and end point and each is assigned a number and a name. Within the Montreal area, a freeway with one number can change names several times and the name changes often have significant functional implications. For instance, the Laurentian Autoroute and the Décarie Expressway are both Autoroute 15. The former is a suburban and rural freeway at ground level while the latter runs in a trench through the central city. Each functionally distinct freeway section is equivalent to a single line in a transit network. The distance from the start of the freeway (chainage) is measured in kilometres and each kilometre is marked by roadside signage. The chainage is used to identify exits but not entrances. The collection of entrances and exits which link multiple facilities (lines) with each other are interchanges and are equivalent to transit stations. In both the road and transit cases, the transfer between lines at stations is associated with a specific impedance.

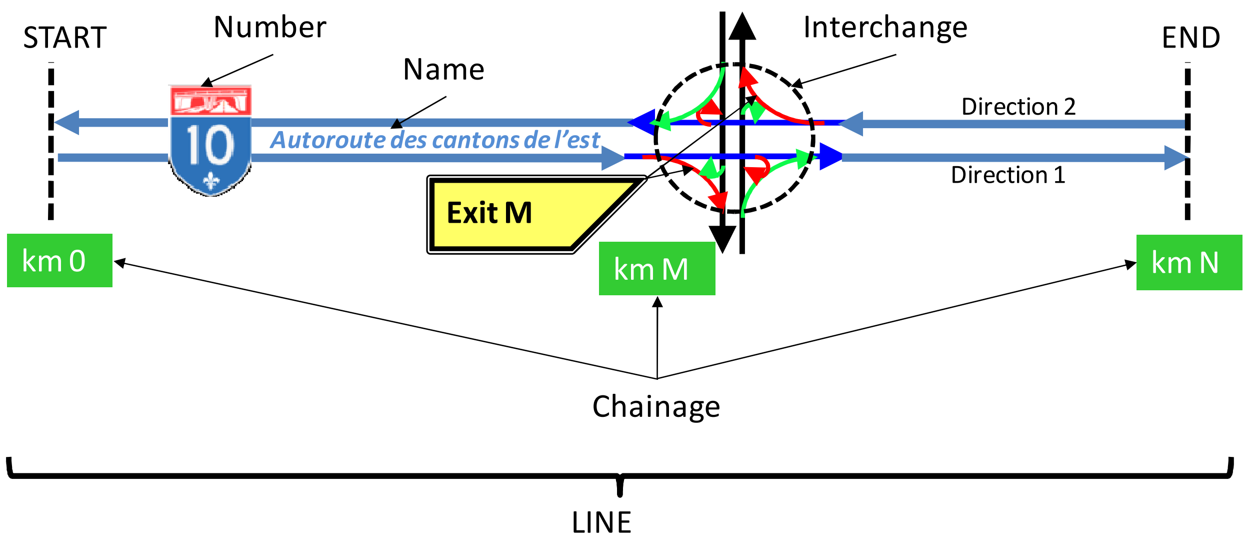


Figure 2.5: The structure of the Québec freeway network

From a modelling perspective, the adoption of the hierarchical approach has three requirements. First, each level in the network hierarchy must be defined; second, the relevant characteristics of each level in the hierarchy must be identified and quantified; third, the importance of penalties associated with the transfer between levels and between lines must be assessed. These requirements are discussed in the next three sections.

2.3.3.2 Hierarchy of the physical infrastructure

The definition of the hierarchical structure is based on the physical structure of the network. As discussed in the Highway Capacity Manual, the components of the physical network can be classified using the paired concepts of mobility and accessibility. Some facilities are designed to provide maximum mobility, others provide maximal accessibility, and still others provide a combination of both. The distinction is important in two respects. First, the configuration of the network composed of mobility-maximizing facilities has a very strong influence on the route choice of drivers. Second, it is the mobility-maximizing facilities which generate the most important redistributive effects. In both cases, infrastructure designed only to provide access is of secondary interest. Indeed, most traffic models representing large urban areas neglect the local road system entirely. Consequently, the construction of a road network model requires a method for determining which roads should be included in the model and which can be safely ignored. This requirement is well-recognized in practice. Equally important, however, is the proper representation of the major infrastructure that *is* included in the model.

Urban road networks are composed of a wide variety of road types, most commonly classified according to the functional designations of freeway, arterial, collector or local. Each functional class is defined by its physical characteristics and associated road geometry. The design criteria for road facilities are based on the mobility-accessibility dichotomy. Facilities which exist solely to provide access are driveways, local roads and ramps. Each type of access facility is associated with a network of limited regional connectivity meaning that the completion of a non-local trip is impossible or highly undesirable. For example, no trip of significant length can be completed using only driveways or only ramps. In some cases, it is possible to travel exclusively on local roads but there almost always exists a faster alternative.

Facilities which exist solely to provide mobility are the highways and freeways. These facilities form networks of limited local connectivity. This means that travel within a small sub-region is impractical. However, the limited local connectivity is the result of a road geometry which controls access and reduces interaction with conflicting traffic. This design allows for high-speed travel.

Arterial and collector roads are generally more difficult to categorize since they provide both access and mobility. Bridges, on the other hand, are easily qualified as mobility infrastructure since they cannot provide access. By definition, a bridge is something that connects two locations separated from each other by some obstacle. The obstacle can be river, a railway, another road or a neighbourhood but in all cases the bridge does not provide access to the obstacle.

2.3.3.3 Characteristics of the road hierarchy levels

In transit assignment models, the choice of line or sequence of lines is determined by attributes of the lines themselves as well as access and transfer characteristics. Line characteristics include comfort, speed and monetary cost. Access, usually accomplished on foot, is considered as a time or distance to be minimized. Transfers may be evaluated based on wait times between vehicles but also based on the amount of walking that is required to make the connection. Central to this model is the concept of impedance or generalized cost which places a weight on different portions of the journey. For example, time spent waiting for a transit vehicle has greater impedance than the time spent in the vehicle itself.

Clear parallels can be drawn with a road network. The lines (major bridges or tunnels, freeways, arterials) can be characterized using the many attributes described in the Highway Capacity Manual, notably the speed limit, the number of lanes, the presence of parking and transit stops, the density of signalized intersections and the degree of coordination between signals. Access is the component of the trip accomplished on the local network which consists of low-speed, low-capacity roads. In the road context, transfers become movements at junctions. Some movements have higher impedance than others. For example, left turns, which must be made during gaps in the oncoming traffic stream, usually require more time and attention than right turns. Similarly, turning movements at an intersection are more demanding than merges and diverges at freeway interchanges. These notions are worth discussing in detail.

2.3.3.3.1 Heavy road infrastructure: bridges and freeways

Mobility-providing freeways and bridges are the focus of the current study. These facilities are sometimes described as "heavy" infrastructure since they play a determining role in regional travel patterns. They also constitute major public works which are almost always associated with a national, as opposed to a regional or municipal, government. The term "heavy" when applied to transportation infrastructure refers to facilities which carry large volumes at high speeds. Within the realm of terrestrial transport, high vehicle volumes and speeds are only possible if the vehicle stream is physically protected from conflicting streams. The physical isolation of conflicting streams is accomplished through grade separation.

Grade separation has important implications for driver behaviour and the distribution of costs and benefits of travel. The minimization of conflict points and the uninterrupted flow regime make heavy road infrastructure especially attractive to drivers. The high average speeds reduce the time spent travelling and the lack of sharp turns as well as minimal braking and accelerating make the driving experience easier. However, the vertical separation between conflicting movements and the imperative of continuous flow on the major facility require the construction of additional roads which serve only to connect the major facility to the rest of the network. As a result, heavy infrastructure can consume a considerable amount of land surface, depending on the particular method used for the grade separation. In addition, the grade separation often removes the road surface from the plane of the surrounding environment. This discontinuity in the urban fabric results either in a trench or an elevated freeway. A trench usually constitutes a physical barrier which reduces the continuity and connectivity of the local and arterial networks. An elevated structure may preserve the surrounding built environment and road network but its utilitarian appearance and purely structural function makes it unwelcome in most communities. For these reasons, the distinction between grade-separated and at-grade intersections must be adequately represented in a traffic model.

A study of the freeway network in Greater Montreal reveals a variety of approaches to grade separation. The most desirable and, due to its prohibitive cost, most uncommon approach is the tunnel. The only major instance in the Montreal area is the Ville-Marie Expressway passing under downtown Montreal. The tunnel removes all interaction of the major infrastructure with the local environment. The infrastructure itself is not visible, and the noise and gases generated by the traffic stream are contained and controlled.

The Décarie Expressway is Montreal's only trench freeway. This facility cuts through a highly urbanized area of the central city. Although the traffic lanes are hidden from view (except in the immediate vicinity of the trench), the local road network is severely disrupted and there is no barrier to protect the surrounding neighbourhoods from the associated noise and air pollution. Moreover, freeway access and egress are provided by a service road whose lanes run parallel to and on both sides of the trench. These lanes intersect all other roads at grade.

The Metropolitain Autoroute (Autoroute 40) is an elevated freeway located in a highly urbanized environment. While the elevated structure separates the traffic lanes from the land surface, the structure itself is built in such a way that the most of the land surface underneath it is unusable. From a land-use perspective, it is exactly equivalent to an at-grade freeway. As with the Décarie Expressway, the Métroplitaine is also accessed via a service road which runs at grade on either side of the freeway. The large distance between the two directions of the service road makes unsignalized intersections impractical. As a result, the facility reduces the connectivity of the local road network.

For modelling purposes, these facilities must be characterized according to quantifiable attributes. In analyses of road transport, the most commonly considered attribute of the built infrastructure is its degree of congestion. User-equilibrium models represent congestion as a high marginal cost of travel on a particular link. This elevated marginal cost causes increments of flow to be assigned to less congested links elsewhere in the network. From the perspective of a driver, congestion is experienced as a low travel speed and a restriction on movement (spacing and lane-changing). Intuitively, a highly congested line or corridor is a less attractive choice than a line or corridor which is uncongested. The fact that congestion does occur, however, indicates that many drivers choose routes even though they are heavily congested. This suggests that drivers may be captive to particular facilities and consequently that traffic congestion alone plays a less important role in the choice of route than is generally believed. This assertion will be tested in the modelling exercise which follows.

Other potentially important line attributes include the free-flow speed, the capacity, and the functional class of the line itself. All these phenomena can be incorporated into a line structure by identifying and ordering the complete sequence of links which form the line. An example is shown in Figure 2.6. The figure is based on output from a conventional static traffic assignment model[2] and illustrates the variation of certain attributes of Autoroute 40 over the section known as the Métropolitaine Expressway. The attributes include a simulated average traffic speed, the specified free-flow speed, the maximum hourly flow rate, the traffic volume for the simulation period and the road capacity. The graphic provides some idea of the experience of a driver using this particular facility.

[2] Modèle de transport de la région de Montréal 2003 (MOTREM 2003), Service de modélisation des systèmes de transport, Québec Ministry of Transport. A thorough description of this elaborate regional transport modelling framework can be found in (Tremblay, 2007) and at:

http://www.mtq.gouv.qc.ca/portal/page/portal/ministere/ministere/recherche_innovation/modelisation_systemes_transport.

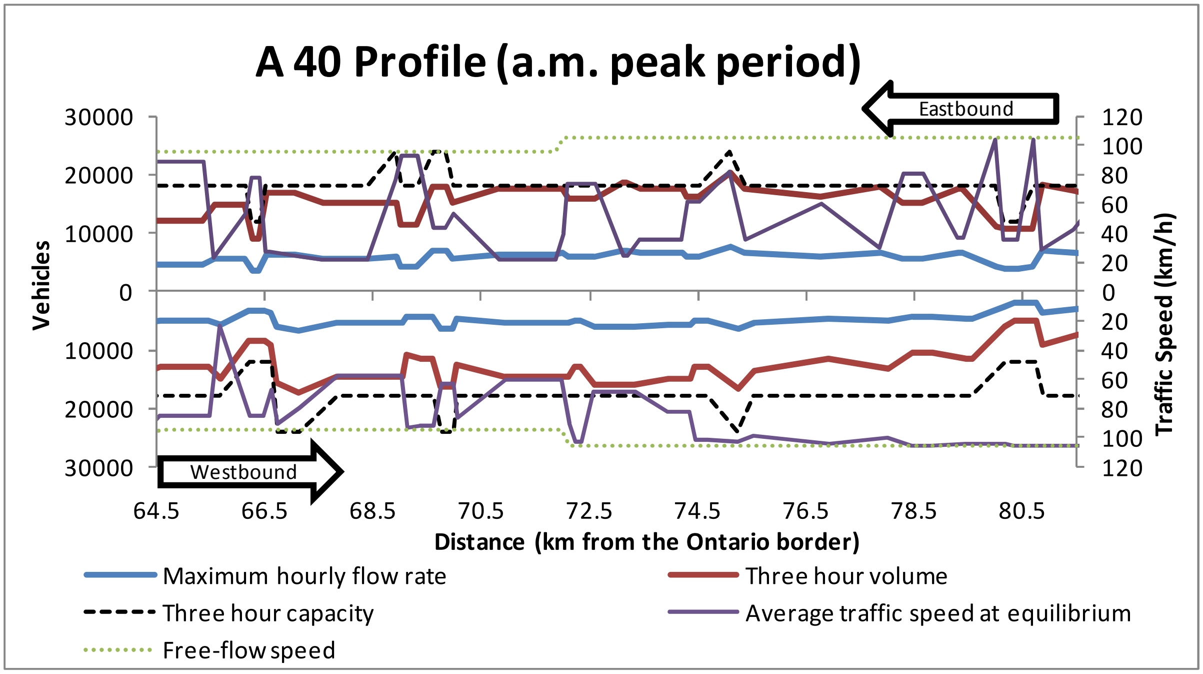


Figure 2.6: The simulated line profile of the Métropolitaine Expressway (Autoroute 40)

2.3.3.3.2 Urban infrastructure: collector and arterial roads

Conflicting flows on urban streets are separated temporally, rather than physically, using traffic signals. The average speed at which traffic moves along an arterial corridor depends on two parameters of the traffic light system: the green time allotted to the corridor and the degree of synchronization. In Montreal, traffic signals are synchronized using coordination networks. A road which passes through multiple networks has multiple coordination schemes which are unlikely to be synchronized with each other. The number of networks encountered could therefore be a proxy indicator of the average speed of traffic along the corridor. A portion of the Montreal street grid and the signal coordination networks is shown in Figure 2.7.

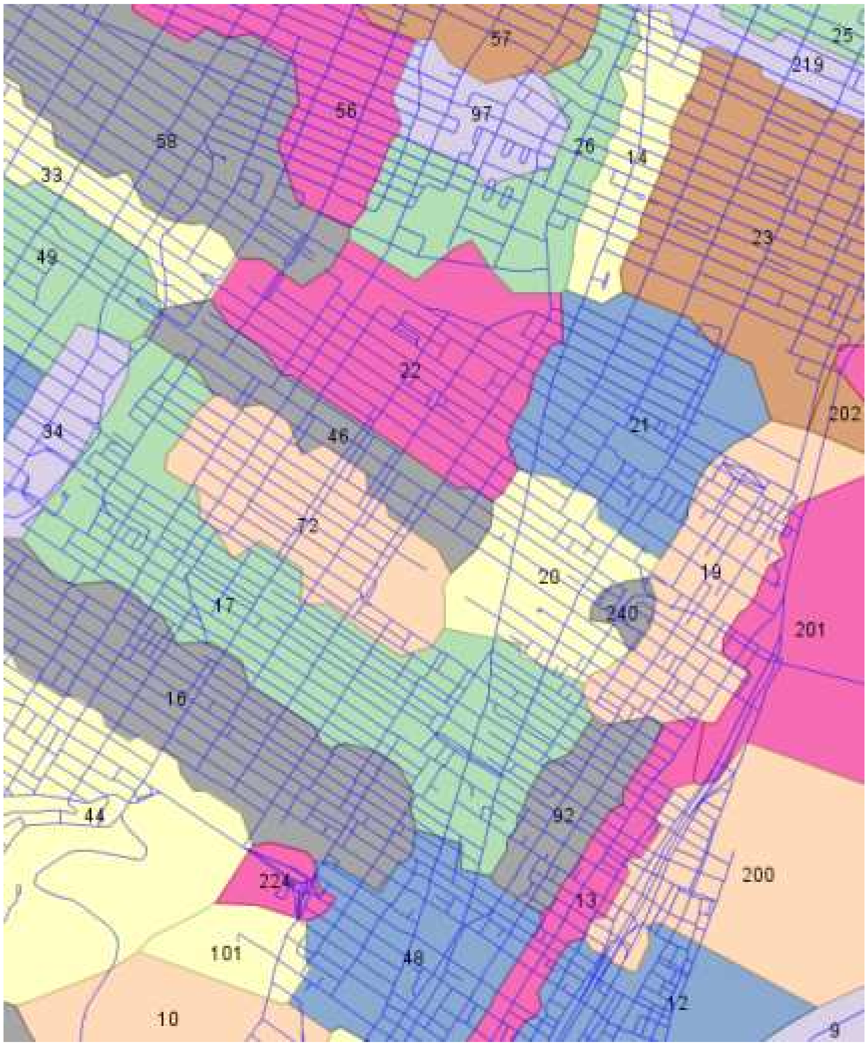


Figure 2.7: Traffic signal coordination networks in Montreal

These considerations are further complicated by the dynamic nature of travel demand. The level-of-service on an urban road network is generally higher at 4 a.m. than at 4 p.m. Here it is possible to draw another parallel with public transit systems: road networks follow a schedule, even if it is not centrally planned. Road network schedules are most notably different from public transit schedules in that they offer poorer service quality at periods of peak demand. The challenge is to assemble and analyse data which represent this temporal evolution and, not at all coincidentally, an estimation of the road network schedule can be obtained with the help of the public transit system.

The performance of the Montreal road network has been monitored for many years by the public transit agency which has an interest in developing good estimates of road travel times at different time periods in order to plan its bus service. The results of this historical compilation are available to the general public in the form of bus schedules. A synthesis of the published schedules can reveal the variation in average network speeds over a specified time interval. Figure 2.8 shows the evolution of average bus speeds on three lines over the course of a typical day based on the published schedule information. It is clear that the buses travel more slowly during the p.m. peak period (15h30-18h30) than during the evening (18h30-1h).

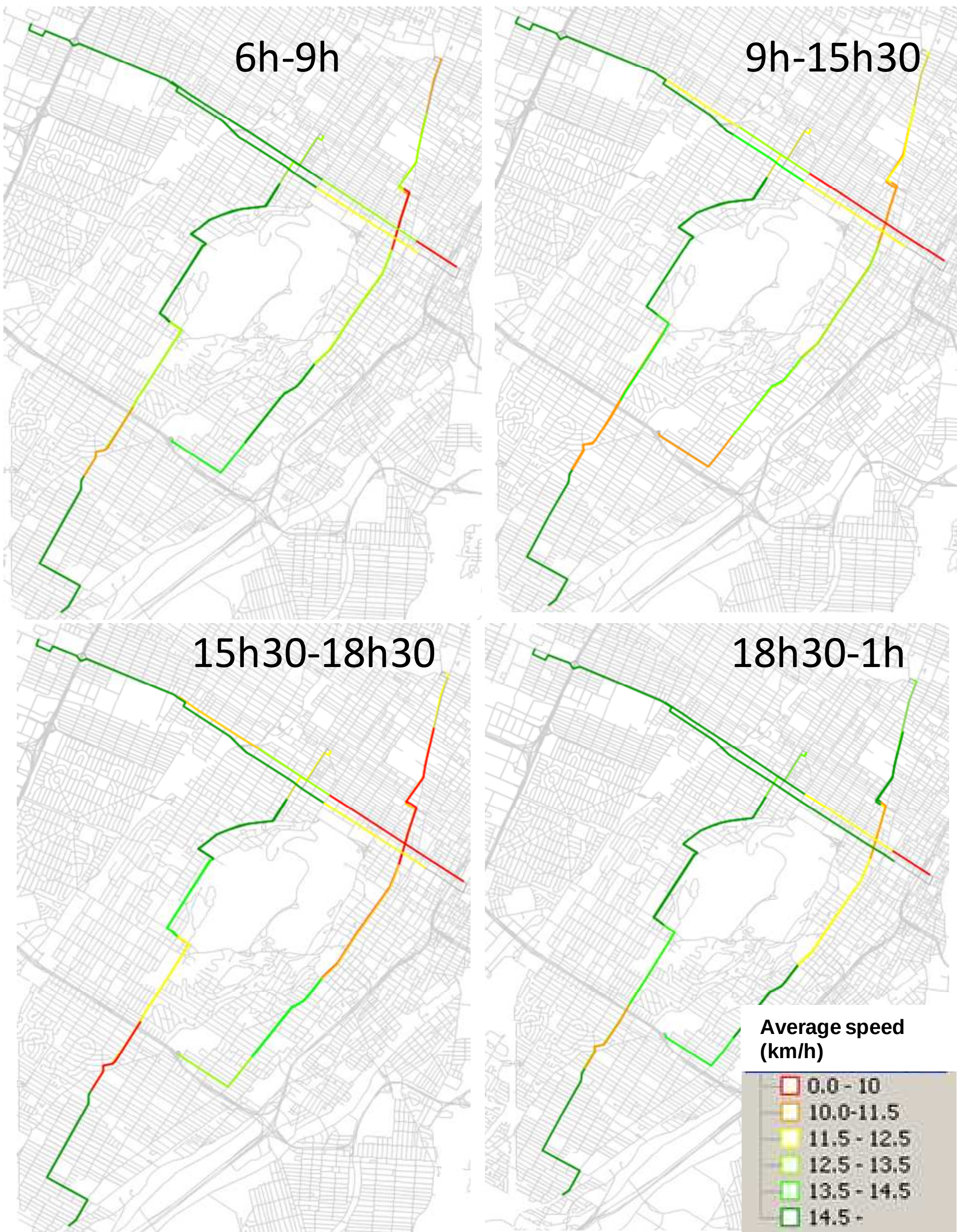


Figure 2.8: The evolution of average commercial speed (km/h) on three bus lines based on published bus schedules.

The compilation of data for a single line (Figure 2.9) shows the detailed variations of speed on the network over the course of an average day. The particular corridor defined by this bus route has a maximum average speed of about twenty kilometres an hour which is achieved very early in the morning. During the day, average speed drops to around 13 km/h and increases above 15 km/h in the evening. There is little difference between the speed profiles of the westbound and eastbound runs. It is not clear how much of the speed variation is attributable to traffic conditions and how much is due to the number of boardings and alightings at bus stops.

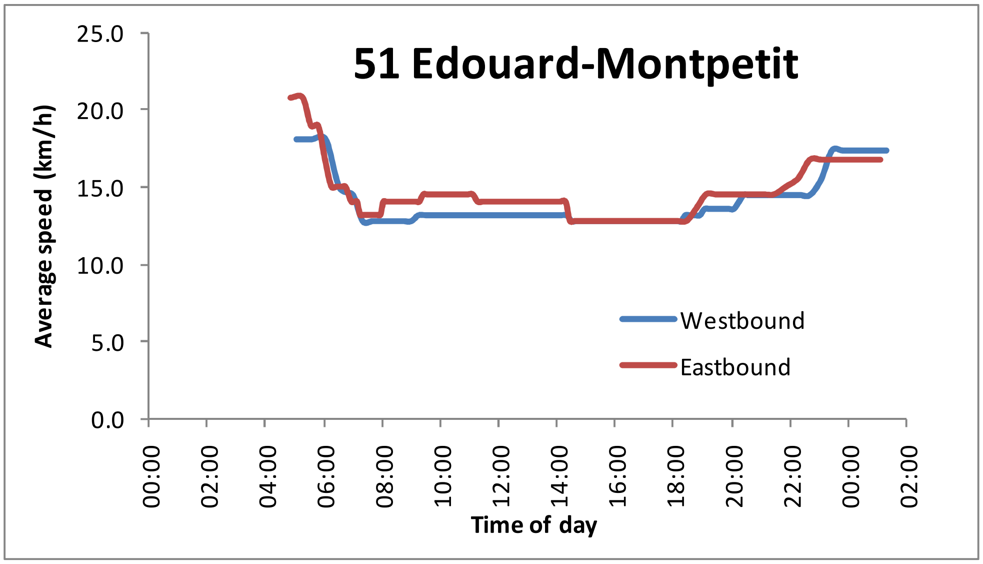


Figure 2.9: Variation in average travel speed for the 51 Edouard-Montpetit bus line

In addition to using planned service information, it is possible, with considerably more computational effort, to compile and synthesize the large quantities of data generated by GPS-equipped public transit vehicles. For about five years, all of the Montreal transit agency's para-transit vehicles were equipped with GPS units which recorded their trajectories. The collected data, consisting of thousands of routes and millions of points, has been successfully used to measure congestion levels on the road network. The para-transit vehicles are better traffic probes than city buses because they are smaller in size and they follow variable itineraries between pre-planned locations, much like private cars. An example of the results of the analysis is shown in Figure 2.10 (taken from (Allard & Grondines, 2007)). This work was used to estimate congested travel times on network links which comprise the Poly network. These travel times have been incorporated into some of the facility choice models described below.

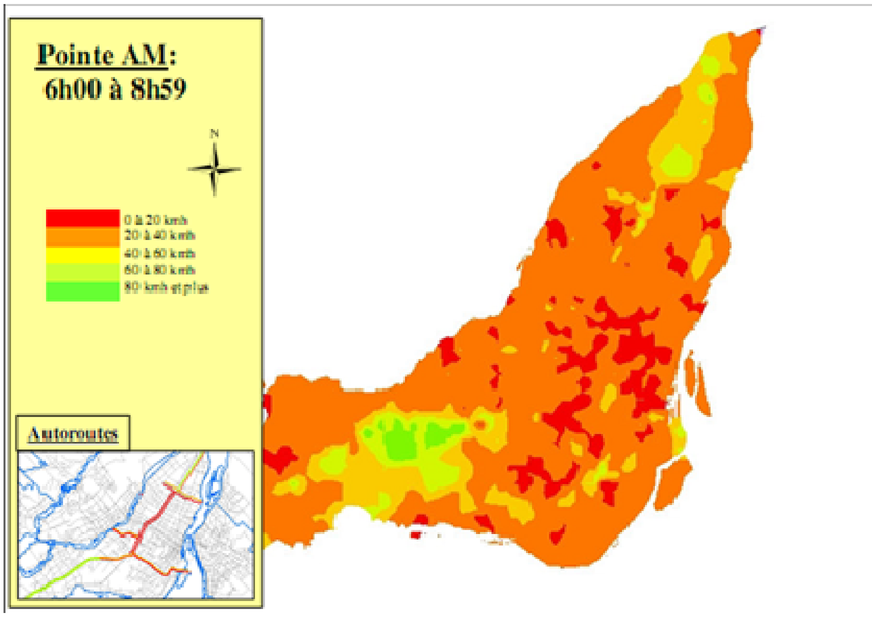


Figure 2.10 : Spatial variation of traffic speeds across the island of Montreal during the morning peak period (Allard & Grondines, 2007).

2.3.3.4 Characteristics of transfer points

In addition to the characteristics of each level in the network hierarchy, the facilities which permit a transfer from one level to another are also important. The transfer from a local road to an arterial or collector road is accomplished through an at grade intersection controlled either by signals or signs. There exists a wide-variety of at-grade intersections subject to evaluation based on safety and convenience. While a detailed study of these facilities is relevant to the planning and modelling of urban roads, it is less important in an analysis of heavy infrastructure (bridges and freeways). The characterisation of transfer points is therefore constrained to transfers between the urban network and the freeway network.

The constraints on the consumption of space in an urban environment influence the design of freeway access points. The typical freeway interchange configuration is based on the “cloverleaf” design which allows access from and to the freeway without stop signs or traffic signals (Figure 2.11). A similar conception is the “directional” interchange (Figure 2.12) whose ramps have much larger curvature radii but they require merges to and diverges from the left lane. In urban areas generally, designs which consume less space are preferred. In the particular case of Montreal, the two most common alternatives are the “roundabout” and the “diamond”. The “roundabout” (Figure 2.13) is constructed using ramps having a small curvature radius where they intersect with the non-freeway road. The non-freeway road itself follows a semi-circular horizontal curve to minimize the angle of intersection with the access ramps. This configuration does not necessarily obviate the need for traffic signals. The diamond (Figure 2.14) is characterized by simple ramps which are controlled by traffic signals at the junctions with the non-freeway road. The operation of “diamond” and “roundabout” interchanges is often facilitated by service roads which run parallel to the freeway lanes. The service roads intersect access ramps and the arterial (and even local) road networks at ground-level. They are often signalized.

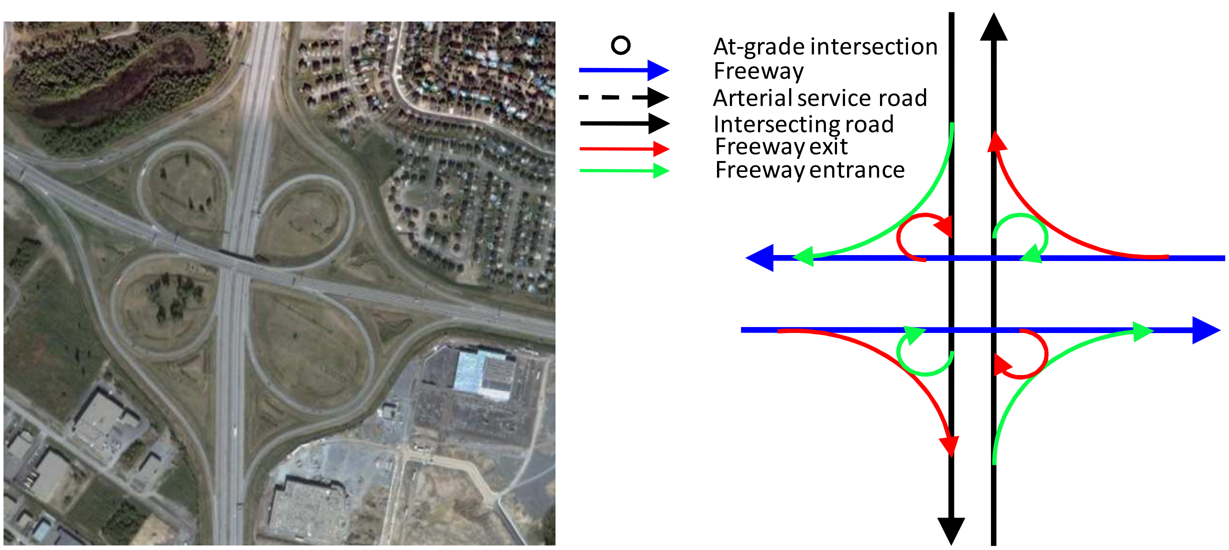


Figure 2.11: A cloverleaf interchange (photo from maps.google.ca)

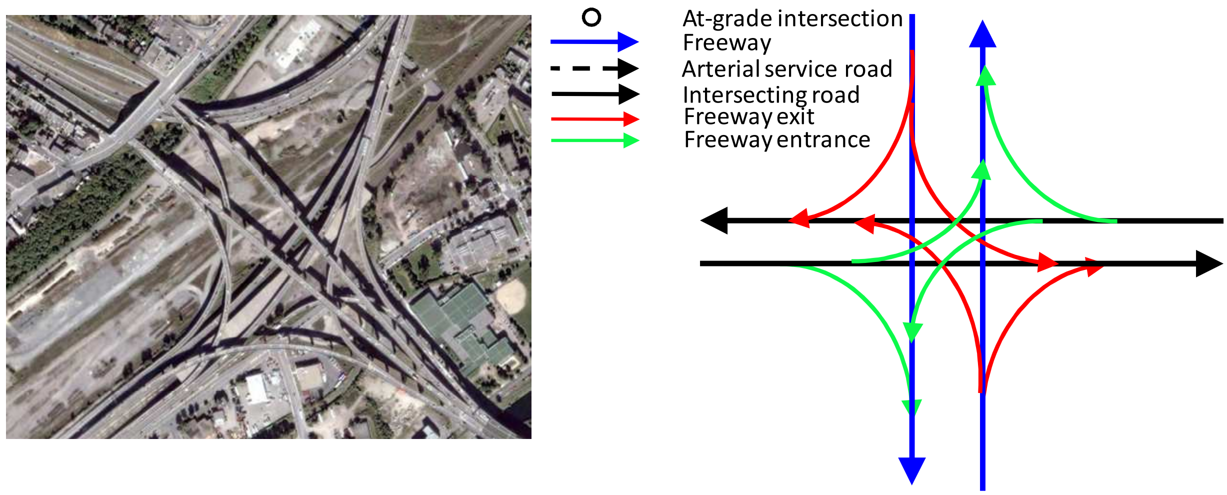


Figure 2.12: A "directional" interchange (photo from maps.google.ca)

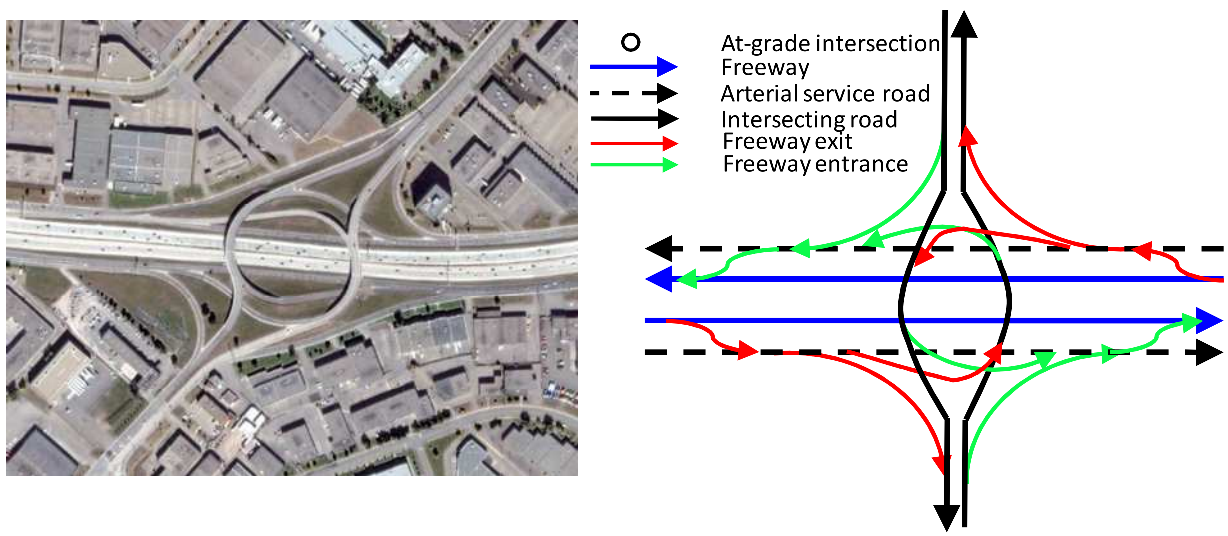


Figure 2.13: A roundabout interchange (photo from maps.google.ca)

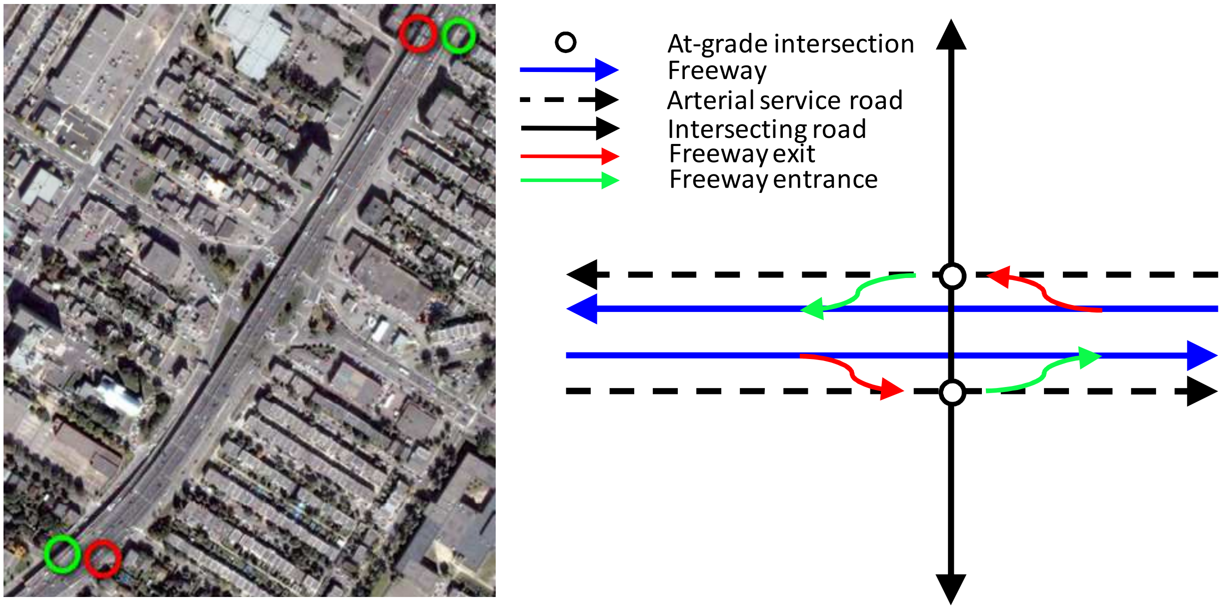


Figure 2.14: A diamond interchange (photo from maps.google.ca)

Each facility type results in a different experience for a given driver. While a perfect cloverleaf interchange allows for an almost effortless transfer between the urban and freeway networks, some of the atypical designs require greater concentration and manoeuvring skill. It seems likely that drivers, depending on their personality and attitudes, will prefer certain interchanges to others if they have a choice. Although the implications for route choice are potentially significant, little is known about the costs or impedances that drivers associate with particular interchanges.

2.3.3.5 Implications for network codification

The important attributes of a particular facility are not always well-represented in traffic models. Road functional class (an attribute not always included in traffic modelling software) provides an indication of which roads are designed for access and which are designed for mobility but the distinction must have some quantitative expression in the model. For example, the posted speed limit on Montréal's Métropolitaine Expressway is 70 km/h. The posted speed limit on Papineau Ave. is also 70 km/h. Both have six lanes. Although the former facility is a freeway and the latter facility is an arterial road, the only functional difference between them is the amount and type of access they provide. The Métropolitaine is only accessible by ramps located at distances of one or two kilometres. Papineau is accessed by arterial roads at signalized intersection.

Note that in a static traffic model it is impossible to find an observed variable that will distinguish the two facilities. Traffic signals are not represented and intersection geometry has no impact at all on link performance. Figure 2.6 shows an example of a case where the distinct appeal of a freeway to drivers is captured by a free-flow speed (100 or 109 km/h) which is much greater than the posted speed limit (70 km/h). Another option is to arbitrarily specify different volume-delay functions for each facility. A dynamic traffic model does incorporate traffic signals and lane geometry but the specific effects of these elements on driver behaviour are not always known. For example, if the traffic lights are perfectly coordinated on Papineau, will a driver adopt behaviour identical to that which he adopts on the Métropolitaine? Do drivers prefer to access a facility using a ramp or a signal? Are some ramp configurations preferable to others? These questions imply that some effort should be devoted to representing the geometry of built infrastructure in a realistic way.

2.4 Exploratory analysis of declared partial path information

The 2003 Montreal travel survey contacted 71,400 households by telephone and questioned each respondent about the trips made by all household members on the day prior to the interview. The survey was conducted over four and a half months from the beginning of September, 2003 until the third week of January, 2004. The expectation is that the collected data can be combined to give a representative image of travel patterns in the Greater Montreal Area during a typical autumn weekday. Daily interviews are conducted regardless of prevailing network conditions on the previous day[3]. This means that non-recurrent events such as road closures and accidents are included in the survey responses. Although these events represent a disruption of the assumed equilibrium state in which the road network exists, their inclusion can be defended based on the belief that the equilibrium is the aggregate result of a “steady-state” system (Holden, 1989). And while specific incidents are certainly deviations from the “average”, unpredictable events in

[3] An important exception was the transit strike of autumn 2003 when the interview schedule was consequently adjusted.

general are a daily occurrence on a busy urban road network and can therefore be considered within the distribution of possible outcomes. On the other hand, in-depth study of drivers' responses to unpredictable road network disruptions would reveal additional aspects of the route-selection process.

The complete database of the 2003 travel survey describes 366,300 trips. Each trip record contains information on the purpose of the trip, the departure time, the mode of transport, the origin and the destination. The age and gender of each traveller is recorded, as well as attributes of the household such as the number of members and the number of vehicles. For trips made by public transit, the bus and metro lines employed are also recorded. An established procedure exists to validate this information. Auto-drivers whose trips implied the use of one or two of the 15 major bridges were asked which bridge they chose. The responses constitute a cross-sectional sample of revealed preference for heavy road infrastructure facilities. They also constitute information which partially describes chosen routes, although a definitive methodology for evaluating these responses has yet to be defined. Therefore, the first stage of the analysis is an evaluation of the coherence and validity of this information for the purposes of regional traffic simulation. The validation process requires the definition of transport objects and the relations between them, developed using the logic of private automobile travel.

2.4.1 Definition of terms and relations

2.4.1.1 Trip

The definition of a trip is the movement from one point to another of a single person for a single purpose. All the trips in this study are completed by auto-drivers. Trips are discrete quantities represented by whole numbers. It is not possible to perform a fraction of a trip.

2.4.1.2 Origin

The origin is the location of the traveller at the moment he begins his trip. It is also the destination of the previous trip. In the present analysis it is defined by a coordinate representing a point on the surface of the Earth.

2.4.1.3 Destination

The destination is the location where the traveller engages in the activity which was the purpose of the trip. In the traditional four-stage model, the origin and destination points are zone centroids. In the totally disaggregate paradigm, they can be any point in space.

2.4.1.4 Facilities

Facilities are the elements of road infrastructure used in the completion of the trip. These include driveways, urban streets, freeways, bridges, tunnels, ramps and intersections. The representation of facilities in a traffic model is determined by the network coding method. By convention all facilities are represented either as links or nodes in the model. Facilities represented by links have entrance and exit points, defined according to the direction of traffic on the facility. In the present analysis, the primary facility of interest is the major bridge.

2.4.1.5 Direction

Because the major bridges of Montreal are network bottlenecks, they are subject to congestion, especially in the prevailing direction of travel demand. Thus the characteristics of a bridge in one direction may be different from its characteristics in the other direction. This difference is especially important during the peak periods. In this context, the most meaningful designation of directions is “inbound” or “outbound” relative to the island of Montreal.

2.4.1.6 Path

A path is a chronologically-ordered sequence of facilities used by a single trip. In a transportation model, each facility is represented by a link or, less commonly, a node. It is generally safer to describe paths using links since ambiguity can arise in cases where two links share the same start and end nodes.

2.4.2 Survey responses

This research aims to discuss paths in the context of information described by a revealed-preference survey. The respondent declares their origin, their destination and the major bridge they used to complete their trip. The direction of travel is inferred from the location of the origin and destination of a particular trip relative to the location of the bridge. At the outset, then, the declared path information is *partial* and consists of a sequence of just 4 points: the origin, the bridge entrance, the bridge exit, the destination.

The island of Montreal is connected to the mainland road network by 15 bridges. Since all but one of these bridges are bi-directional, they represent 29 choice alternatives although the regional geography ensures that the choice set (for a trip involving exactly one bridge) will never contain more than 8 options. Table 2.4 is a summary of all bridge declarations in the 2003 travel survey. The bridge volumes in the table are calculated as the sum of the expansion factors of all observed trips. The expansion factor for a given trip is calculated based on the demographic characteristics of the traveller. Each age-sex cohort is weighted according to the demographic distribution observed in the 2001 national census. Although the responses have not been validated, the resulting traffic volume estimates are generally consistent with common knowledge of Montreal bridge infrastructure with the Laval and South Shore screenlines being the busiest. The bridges carrying freeways experience higher volumes than the bridges which carry smaller roads.

Table 2.4: Bridge volumes derived from the declarations of the 2003 travel survey

Declared bridge			**Period**			
Name	**Number**	**Screenline**	**A.M. Peak (6:00-9:00)**	**P.M. Peak (15:30-18:30)**	**Off-peak**	**TOTAL**
Champlain	1301	South Shore	16861	17070	31406	65338
Victoria	1302		6841	6629	8731	22202
Jacques-Cartier	1303		17641	18066	30120	65827
L.-H.-Lafontaine	1304		16057	17946	27797	61799
Mercier	1501		11018	12064	20743	43825
Galipeault	1503	West	5982	6471	8981	21434
Île-aux-Tourtes	1504		12133	12280	18093	42506
Viau	1401	Laval	13282	13125	18067	44474
Papineau	1402		5427	5316	6608	17352
Pie-IX	1403		23086	23698	37688	84472
Lachapelle	1404		20628	21139	32907	74674
Médéric-Martin	1405		6911	6765	9587	23263
Louis-Bisson	1406		8689	9251	11361	29301
Le Gardeur	1601	East	3963	4460	4567	12990
Charles-de-Gaulle	1602		14875	14196	23031	52102
	TOTAL		**183395**	**188476**	**289687**	**661559**

The distribution of this surveyed demand by time of day, screenline and direction is shown in Figure 2.15. The figure also illustrates the results of bridge traffic counts performed by the Québec Ministry of Transport (MTQ) coincident with the conduction of the travel survey in autumn 2003. In the figure, the count data are segregated by direction and time of-day but not by screenline. Since the travel survey captures only non-commercial travel, passenger car must be isolated in the traffic count data. At 12 of the 29 counting stations, the collected traffic volumes were stratified by vehicle class. At locations where only the total volume of traffic was reported, an adjustment factor was applied to estimate the volume of passenger cars. This adjustment factor was constructed using the data from bridges where volumes by vehicle class were measured. The figure shows that the volumes traffic volumes obtained using the expansion factors of surveyed trips correspond well with traffic counts, particularly during the morning peak period between 6 and 9 a.m. During this time interval, the traffic volumes inbound (toward Montreal) are better represented than the volumes exiting the island (outbound). The latter are consistently under-represented in the survey. During the p.m. peak period, the survey corresponds better with outbound rather than inbound volumes. In addition, the figure shows that the use of personal expansion factors to weight travel survey trip records produces a systematic underestimation of bridge traffic volumes. Compared to the morning peak period this underestimation is considerably more pronounced in the p.m. peak period and during off-peak periods. This phenomenon may be attributed to the underreporting of complex trip chains and survey respondents' incomplete knowledge of the travel patterns of the other members of their household. The increased presence of external traffic as the day progresses may also play a role. More generally, the roadside counts are collected over three or four days at each location and capture all vehicles using the bridges. The travel survey was performed over nearly five months and describes the travel behaviour of households only. Discrepancies between the bridge volumes calculated using the survey and the directly observed vehicle volumes are therefore to be expected.

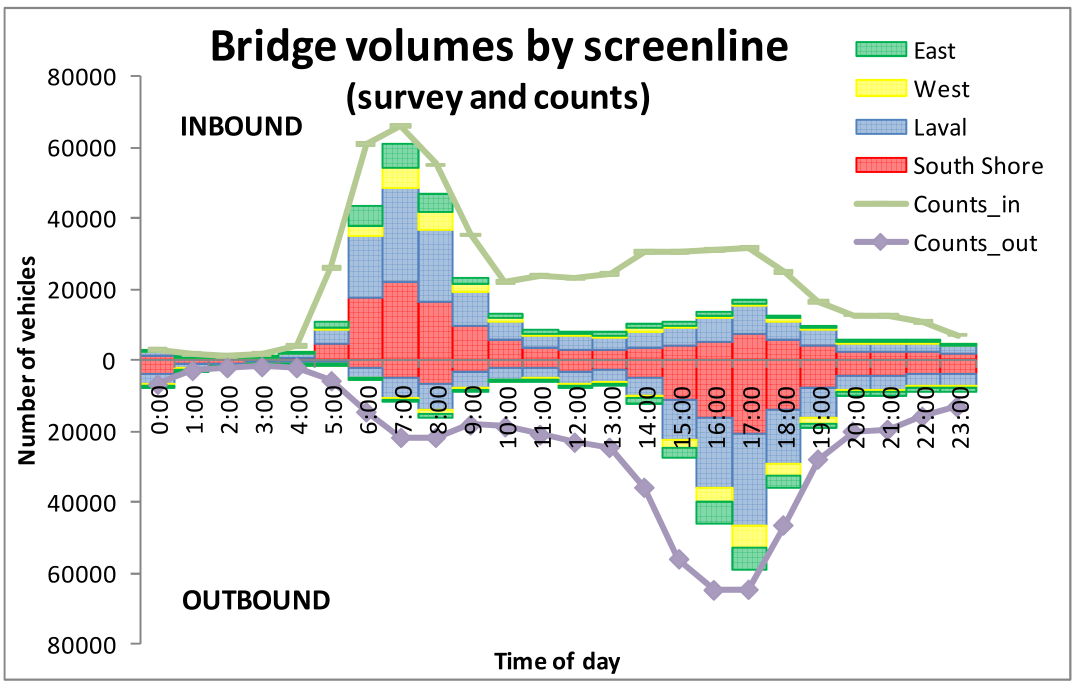


Figure 2.15: Distribution of major bridge declarations and bridge traffic counts by screenline and departure time

The aggregate results are complicated by trips which use multiple bridges (two is the maximum possible for all non-circuitous trips). Multiple bridge declarations are rare however, accounting for 804 of 33,194 observations (2.4%). For the a.m. peak period, there are 185 two-bridge declarations (2.0%). The joint-distribution of paired bridge responses shown in Figure 2.16 demonstrates the overall coherence of the observed choices. For the most part, the sub-matrices representing intra-screenline movement are empty. This means there are very few travellers who enter the island and leave the island using the same screenline. The more frequent bridge pairings confirm some intuitions about bridge usage patterns. For example, the most frequently used second crossing is the Lafontaine Tunnel (1304). Since this facility is located far from the central city and is directly connected to several freeways, it is an important corridor for travel between suburban communities. It acts in concert with the nearby Pont Charles-de-Gaulle (1602) which provides access to the island from the east and the communities north of Laval. There is also a strong interaction with the Pont Pie-IX which is the easternmost link to Laval and shares Autoroute 25 with the Lafontaine Tunnel. Trips using two bridges are excluded from the subsequent analysis. Although they account for a small proportion of total bridge traffic, these trips may influence the level of service on certain crossings. An evaluation of the magnitude of this effect could be the subject of future research.

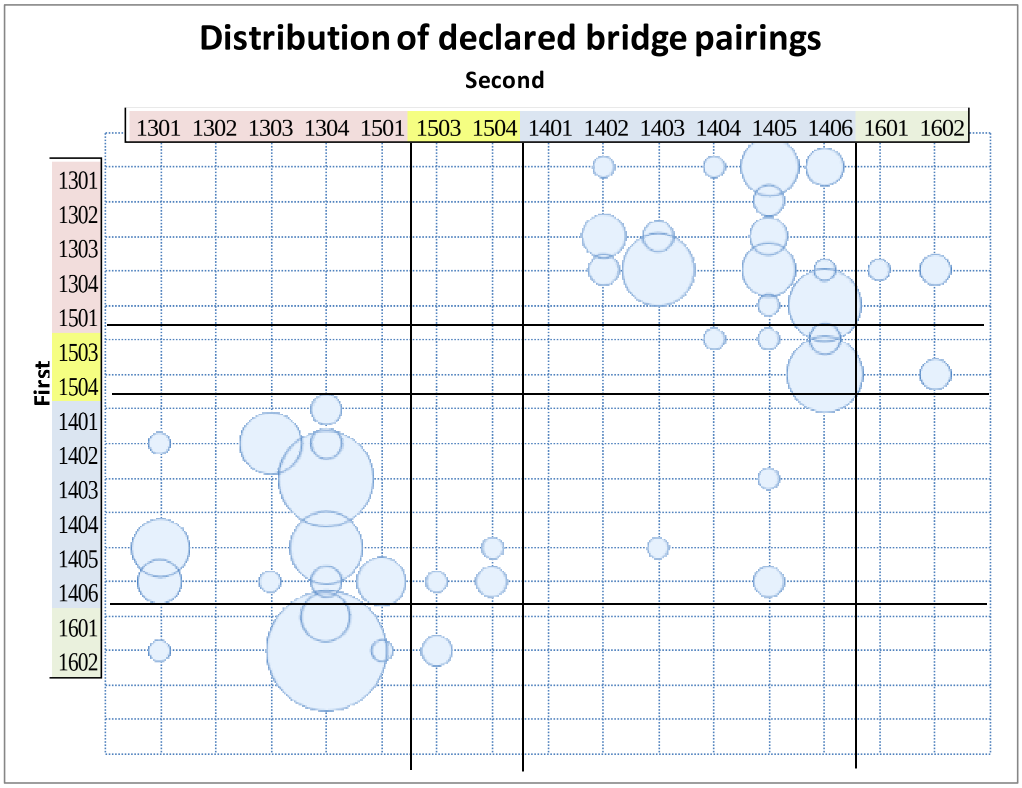


Figure 2.16: Distribution of bridge combinations for trips using two bridges during the morning peak period

2.4.3 Construction of a valid sample

Although the information in the survey paints picture of driver behaviour which is plausible at the aggregate level, legitimate questions can be raised about the accuracy and credibility of individual declarations bridge choice. It is possible that travellers confuse bridges during the interview, especially if they are replying on behalf of another household member. It is also possible they are unable to identify the bridge they used. While the five South Shore bridges are well-known throughout the region, relatively few inhabitants can correctly name all six of the bridges linking Montreal to Laval or the bridges of the East and West screenlines. Additional problems can be attributed to improper geocoding of origins or destinations. It is therefore essential that a method be found for validating these partial descriptions of trip itineraries.

The validation of bridge declarations proposed here consists of associating the declared responses with basic information describing the built infrastructure. This association process is based on a hypothesis of driver behaviour – a simple traffic assignment model. The validation procedure is restricted to trips beginning within the a.m. peak period which by definition starts at 6:00 and ends at 8:59. There are several reasons for adopting this approach. First, a smaller sample facilitates the elaboration and execution of an experimental methodology which could subsequently be applied to a larger set of observations. Second, it is important that a period of peak demand be selected if issues relating to traffic congestion are to be explored. The typical morning rush hour represented by the survey data constitutes a snapshot of the equilibrium state under congested conditions. Finally, there is some evidence to suggest that the survey provides a more complete description of travel demand in the morning peak period than in the afternoon (see section 2.4.2).

2.4.3.1 The validation network

The validation process requires a network on which the declared responses can be simulated. In this case, the Poly network (see section 2.3.2) was used. A portion of this network is shown in Figure 2.17. In this network, as with all the others which are used in subsequent analyses, there is no system of zones, no centroids, and no artificial connecting links. All trip origins and destinations are geocoded to the nearest metre. These trip endpoints are associated to network nodes using a nearest-neighbour method.

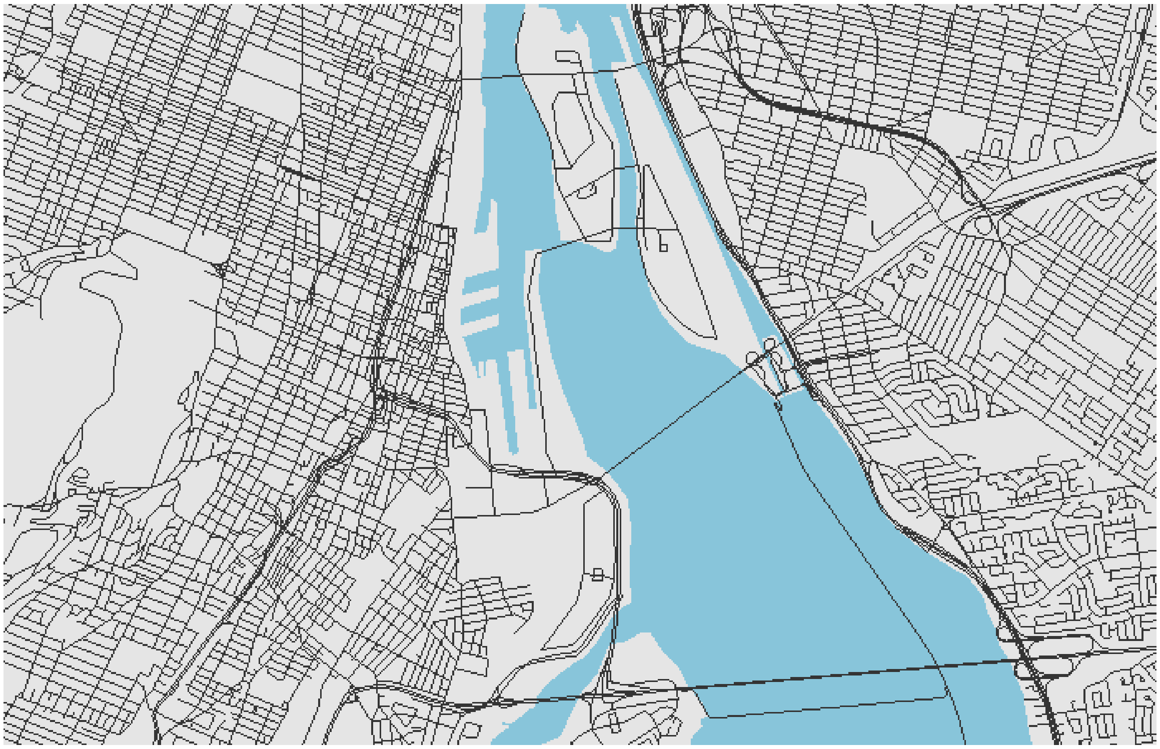


Figure 2.17: The detailed geomatic network for the assignment of trips

2.4.3.2 Representing demand: a disaggregate analysis framework

The majority of traffic assignment models used in practice rely on some sort of zone system and assign demand from centroid to centroid. Under the user-equilibrium hypothesis, this inter-zonal demand is distributed among alternative paths having almost equal travel times. Various methods have been devised for disaggregating the loading process in time through the construction of time-sliced matrices and dynamic loading algorithms. Spatial disaggregation, however, is anything but standard practice because of a historical lack of precisely geocoded trip data and the difficulties involved in achieving a user-equilibrium when the modelling units are individual vehicles rather than packets of flow.

The method proposed here assigns trips to the network from a list. Because they are extracted from the Montreal Travel Survey each trip has many attributes, but the most important ones from the perspective of traffic assignment are the point of origin, the point of destination, and the time of departure. The assignment model is merely the construction of a new trip attribute called the itinerary which is a sequential list of the links used to complete the trip. Furthermore, the structure of the itinerary data represents the cognitive process of the modelled agent. Such representations of knowledge are extremely common in traveller-information systems, an example of which is the service provided by Google Maps (Figure 2.18). The origin and destination are chosen as points (addresses or intersections) and the path is displayed as well as described as set of directions. Note that the directions constitute a list of intersections where a turn is required. This node-movement representation of a trip is what ought to be preserved as the output of any meaningful traffic assignment or route choice model since it mirrors the cognitive structure used by drivers as they navigate through the network. It also provides complete information on the physical infrastructure employed by travellers.

Figure 2.18: Example of routing output from Google maps (maps.google.ca)

One of the limitations of using a travel survey to represent demand is that the survey contains only a sample of the total population. The 2003 Montreal travel survey sampled roughly 4.5% of the region's households but its sampling rate of trips is generally lower, as suggested by the comparison of expanded survey trips to observed bridge volumes shown in Figure 2.15. An expansion factor is estimated for each validated trip in the survey but, since it is based solely on demographics, it does not necessarily generate a legitimate estimate of traffic flows at particular locations or even of total automobile travel demand on the network generally. Also, the conventional use of the survey expansion factors introduces an important distortion during trip simulation because, instead of one person travelling from point A to point B along route X, there are 25 people exhibiting identical behaviour. What is required is a legitimate method for disaggregating – in space and in time – individual survey records after they have been weighted. Since the modelling exercise undertaken in this thesis is focused on the faithful reproduction of the observed behaviour of individual drivers, the analysis uses unexpanded (non-weighted) trip records. The faithful reproduction of observed link flows validated by roadside counts constitutes a different problem. Nonetheless, a representation of the full population of vehicles is necessary when simulating traffic congestion. This particular issue is discussed further in section 2.5.2.1.

2.4.3.3 Application of two all-or-nothing assignments

For validation purposes, it is assumed that drivers are optimists in that they believe traffic congestion will be minimal and therefore has no influence on their choice of route. They are also selfish optimizers in that they choose the route which minimizes their own travel time. An all-or-nothing shortest path assignment is an appropriate algorithm for representing these hypotheses. Trips are assigned to the network sequentially. The path information for each assigned trip is retained as a final result of the simulation process.

Two simulations are performed. The first one is an all-or-nothing assignment of trips to the complete network. The second simulation is an all-or-nothing assignment where only the bridge declared by the survey respondent is available. For each survey observation the attributes of the shortest path are compared to the attributes of the declared path. An initial comparison of the travel time attribute for the 9,290 bridge choice observations in the a.m. peak period is shown in Figure 2.19. The shortest path travel time is on the vertical axis and the minimum travel time for the itinerary incorporating the declared bridge is on the horizontal axis. Observations which fall far from the axis of symmetry are considered suspect.

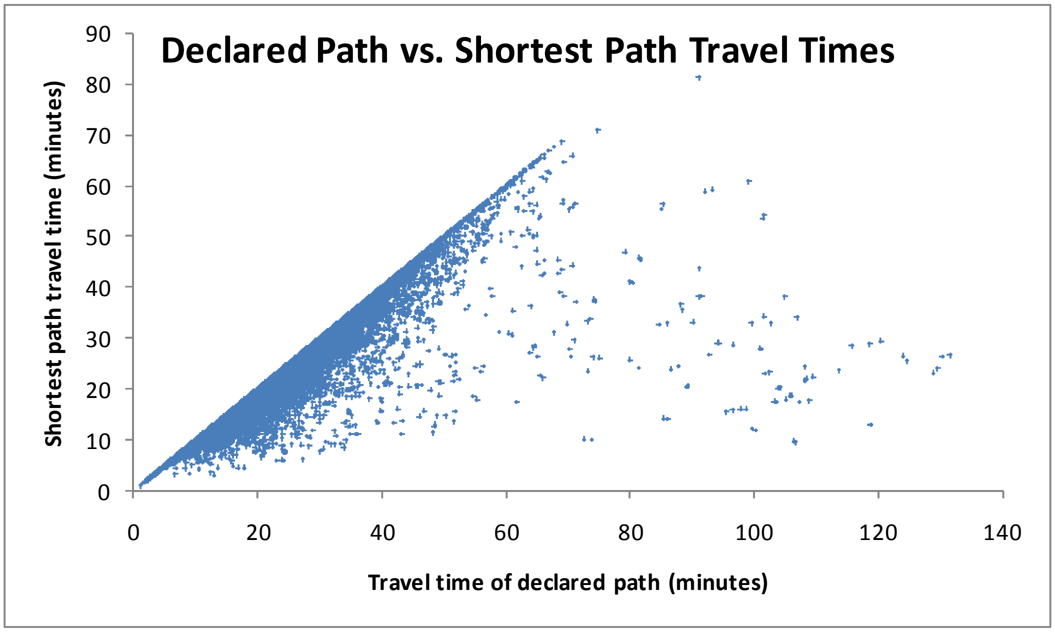


Figure 2.19: Comparison of simulated travel times of declared routes and shortest-path routes.

If the travel time on the declared path exceeded the travel time on the shortest path by more than 15 minutes, the observation was removed from the dataset. An absolute, as opposed to a relative, threshold was chosen based on the assumption that drivers do not distinguish between small travel time differences, even if these differences represent a significant proportion of the total travel time. Observations which fell very close to the 15 minute threshold were subject to a visual inspection of their declared itineraries. In some cases, this exercise revealed obvious inconsistencies in the spatial distribution of trip origins, destinations and the declared bridge. These observations were removed from the dataset. Additional responses were removed because the declared mode of travel was not auto-drive. For the a.m. peak period (6 a.m. to 8:59 a.m.), the final sample contained 8,583 observations.

2.4.3.4 Description of the validated sample

The trip records contained in the validated subsample were generated by interviews performed between the 27th of August, 2003 and the 22nd of January, 2004. Although the survey is designed to represent travel behaviour for a typical weekday in the fall, a strike by maintenance workers at Montreal's public transit agency (the Société de transport de Montréal or STM) from the 18th to the 23rd of November caused a disruption in travel patterns during that period. As a result, the decision was made to extend the interview period into the winter of 2004. No interviews were conducted between the 11th of December and the 6th of January because of the important changes in travel patterns that occur around the holidays. The temporal distribution of trips in the validated sub-sample over the surveyed period is shown in Figure 2.20. The subsample includes three days during the transit strike and three weeks in January. The volume of observations during the strike is noticeably lower than at other times. Since “essential” transit service was maintained during rush hours, off-Island transit networks were unaffected, and the number of observations made during the strike represents a small proportion of the sub-sample, it is unlikely that the inclusion of trips made during the strike introduces an important bias into the representation of bridge usage patterns. The gradual increase in the daily volume of trips apparent over the course of September is attributable to the initial low productivity of the call-centre and not to a dramatic increase in bridge usage over a period of a few weeks.

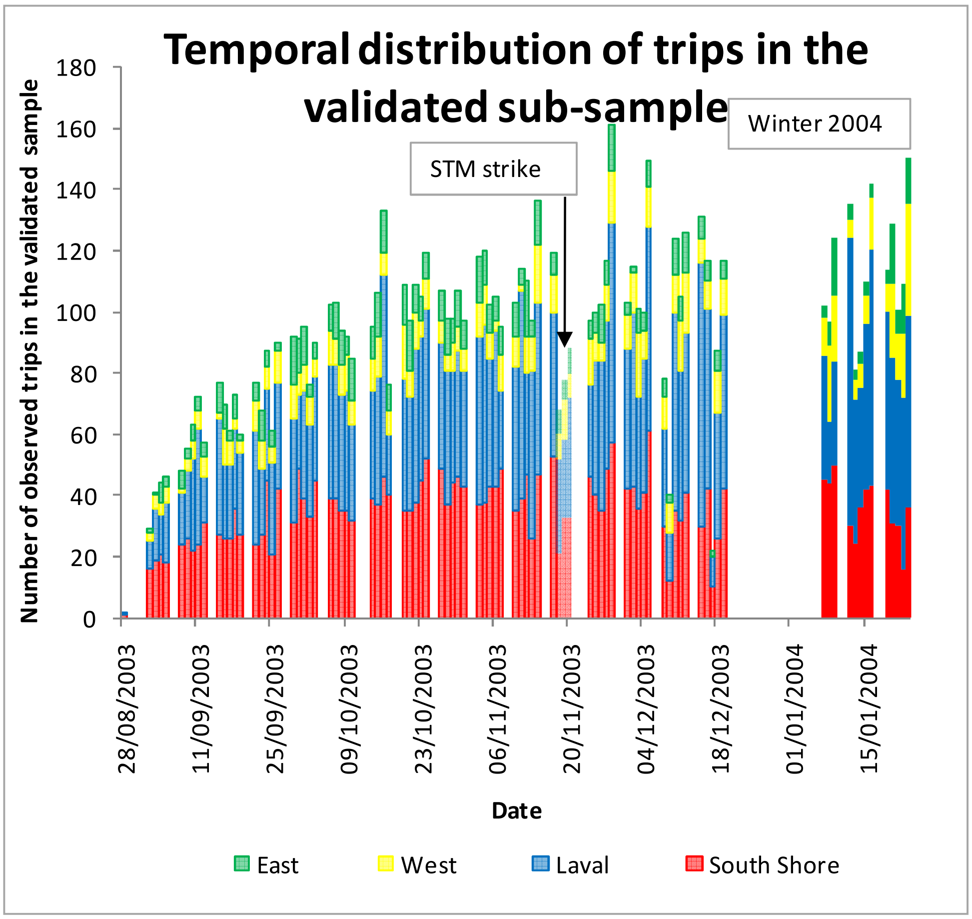


Figure 2.20: Temporal distribution of validated trips by screenline

Because travel patterns may vary over the course of a typical week, it is important to verify the stability of bridge usage patterns for each weekday. Figure 2.21 compiles the number of trips in the validated subsample by screenline and by weekday. The figure shows that the distribution of trips over the four screenlines hardly changes from one day to the next. A slight increase in the total number of bridge-using trips is apparent on Thursdays and Fridays. An examination of trip motives revealed no change in the frequency of particular types of trips, suggesting that the increased number of observations is due to an improved quantity and quality of survey responses at the end of the week.

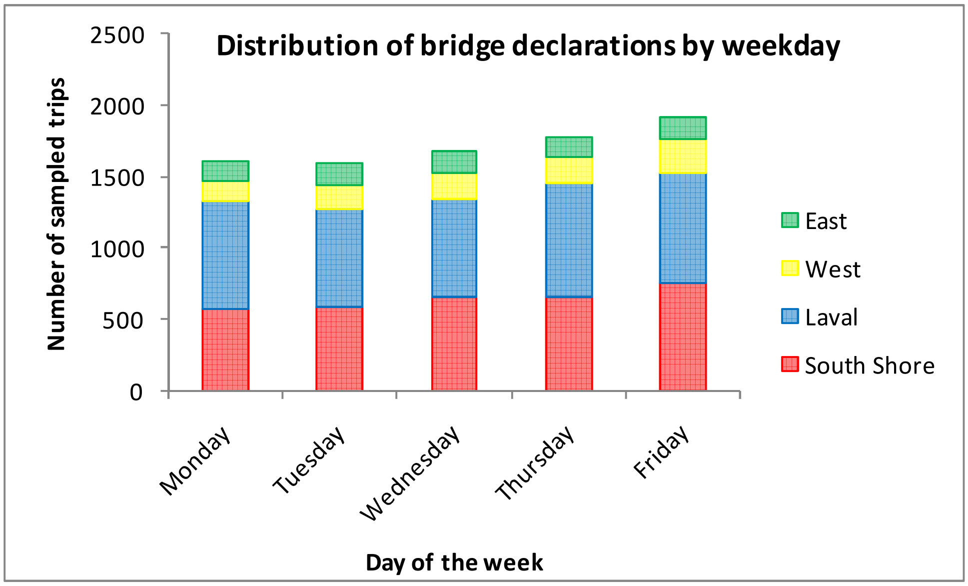


Figure 2.21: Distribution of observed trips over the days of the week by screenline

The validated survey records can also be examined by bridge. Table 2.5 presents some summary statistics of the responses for each of the 15 bridges over the 91 days during which bridge choices were recorded. Of particular interest are the bridges that carry non-freeway roads (1302, 1401, 1404 and 1601). Compared to the other eleven bridges, these bridges appear less frequently in the sample. They have a lower average number of responses and a higher number of days without any observations. This result is to be expected since arterial roads normally carry much lower volumes than freeways. There is, however, an interesting variation in response attributes among these four bridges. At one extreme, the Victoria (1302) bridge declarations appear to be the most reliable based on the average number of responses per day and the coefficient of variation (the standard deviation divided by the mean). At the other extreme, the

Le Gardeur bridge (1601) is the most poorly represented with 30 days of non-response and a very high coefficient of variation.

Table 2.5: Summary of survey responses by declared bridge over 91 days of interviews

Bridge	Total responses	Days of no response	Max responses	Avg response per day	Std. dev.	Coef. of variation
1301	778	1	22	8.55	4.21	0.493
1302	360	5	9	3.96	2.19	0.553
1303	828	2	20	9.10	3.87	0.425
1304	714	1	18	7.85	3.53	0.450
1501	549	4	14	6.03	2.94	0.487
1401	243	12	12	2.67	1.99	0.747
1402	560	4	14	6.15	3.35	0.544
1403	654	2	25	7.19	4.30	0.599
1404	213	11	6	2.34	1.56	0.665
1405	1018	1	25	11.19	5.21	0.466
1406	1007	1	26	11.07	6.05	0.547
1503	408	3	22	4.48	3.64	0.812
1504	511	4	16	5.62	3.07	0.546
1601	135	30	6	1.48	1.49	1.001
1602	605	1	16	6.65	3.74	0.562

The breakdown of observations by bridge, screenline and direction for the validated data set is shown in Table 2.6. The comparative totals are consistent with the distribution of traffic counts[4] among the facilities, which are also shown in the table. The two leftmost columns are the expansion factors which would need to be applied to the survey responses for them to match the observed traffic volumes. Since the survey samples around 5% of the region's households, a naïve trip expansion factor should be around 20. For most bridges, the inbound factors are not far from this value although they are consistently above 20, suggesting a generalized under-representation of bridge demand in the travel survey. This finding is consistent with the analysis of counts over the course of an average day (see section 2.4.2). Large discrepancies are apparent on the Lachapelle and Le Gardeur bridges. The much bigger expansion factors on outbound flows indicate that the survey systematically underestimates demand in the non-peak direction.

[4] These data were collected by the Québec Ministry of Transport during the fall of 2003, which corresponds to the time at which the survey was conducted.

The four leftmost columns of Table 2.6 show the surveyed trips and counts normalized by directional (inbound/outbound) screenline. These percentages allow a comparison of the distribution of traffic across the facilities which form each screenline. For example, 23.6% of all surveyed trips crossing the South Shore screenline in the inbound direction used the Champlain Bridge. Meanwhile, 23.9% of inbound vehicle traffic observed on the South Shore screenline used the Champlain Bridge.

Table 2.6: Validated bridge declarations for the a.m. peak period

		Surveyed Trips			Traffic counts		Expansion factors		Trips (screenline normalized)		Counts (screenline normalized)	
Bridge	**Screenline**	**Out**	**In**	**TOTAL**	**Out**	**In**	**Out**	**In**	**Out**	**In**	**Out**	**In**
1301	South Shore	156	622	778	6999	15215	44.9	24.5	26.2%	23.6%	26.0%	23.9%
1302		-	360	360	-	8095	-	22.5	0.0%	13.7%	0.0%	12.7%
1303		124	704	828	5505	15016	44.4	21.3	20.8%	26.7%	20.5%	23.6%
1304		220	494	714	10107	13850	45.9	28.0	37.0%	18.8%	37.6%	21.7%
1501		95	454	549	4260	11517	44.8	25.4	16.0%	17.2%	15.9%	18.1%
South Shore TOTAL		**595**	**2634**	**3229**	**26871**	**63693**			**100%**	**100%**	**100%**	**100%**
1503	West	32	376	408	2299	9099	71.8	24.2	38.1%	45.0%	38.3%	42.3%
1504		52	459	511	3709	12420	71.3	27.1	61.9%	55.0%	61.7%	57.7%
West TOTAL		**84**	**835**	**919**	**6008**	**21519**			**100%**	**100%**	**100%**	**100%**
1401	Laval	37	206	243	2069	5643	55.9	27.4	5.3%	6.9%	6.3%	6.7%
1402		112	448	560	4345	8932	38.8	19.9	16.1%	14.9%	13.2%	10.6%
1403		138	516	654	7181	14094	52.0	27.3	19.8%	17.2%	21.8%	16.7%
1404		31	182	213	1701	8034	54.9	44.1	4.5%	6.1%	5.2%	9.5%
1405		210	808	1018	9406	23281	44.8	28.8	30.2%	26.9%	28.6%	27.6%
1406		168	839	1007	8179	24241	48.7	28.9	24.1%	28.0%	24.9%	28.8%
Laval TOTAL		**696**	**2999**	**3695**	**32881**	**84225**			**100%**	**100%**	**100%**	**100%**
1601	East	12	123	135	948	6076	79.0	49.4	17.6%	18.3%	13.0%	24.7%
1602		56	549	605	6327	18556	113.0	33.8	82.4%	81.7%	87.0%	75.3%
East TOTAL		**68**	**672**	**740**	**7275**	**24632**			**100%**	**100%**	**100%**	**100%**
TOTAL		**1443**	**7140**	**8583**	**73035**	**194069**	**50.6**	**27.2**				

The market shares of each facility measured using counts are very strongly correlated the market shares derived from the survey. A plot of normalized counts vs. normalized survey trips illustrates the strength of the correlation (Figure 2.22). A linear regression weighted by the number of survey observations yields a determination coefficient (r-squared) of 98%. This finding strongly suggests that the validated survey information is an accurate representation of driver bridge choice.

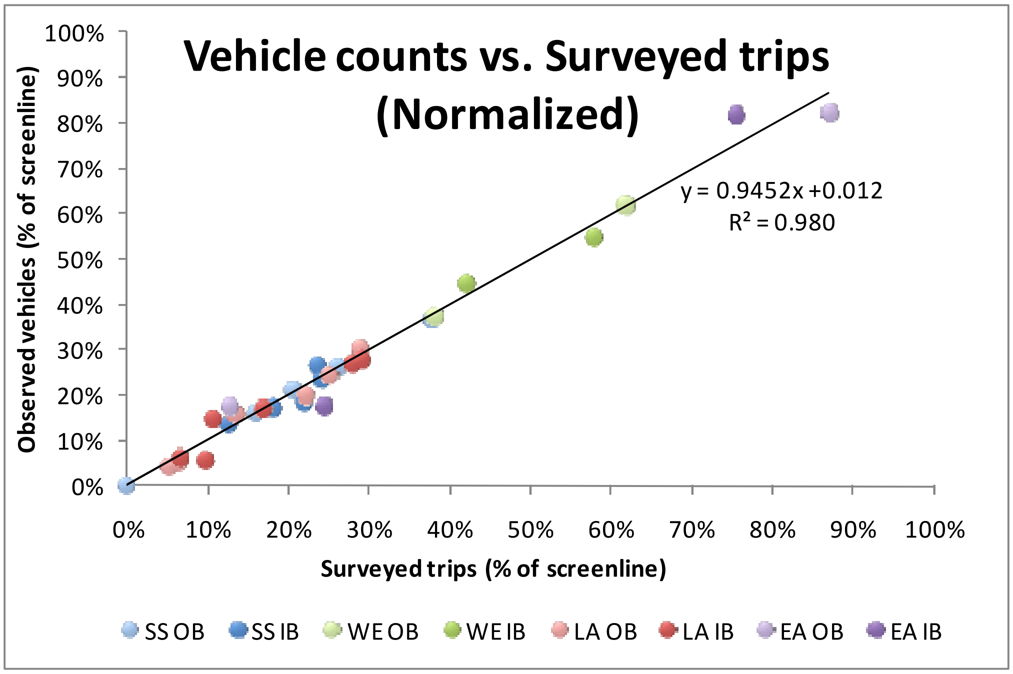
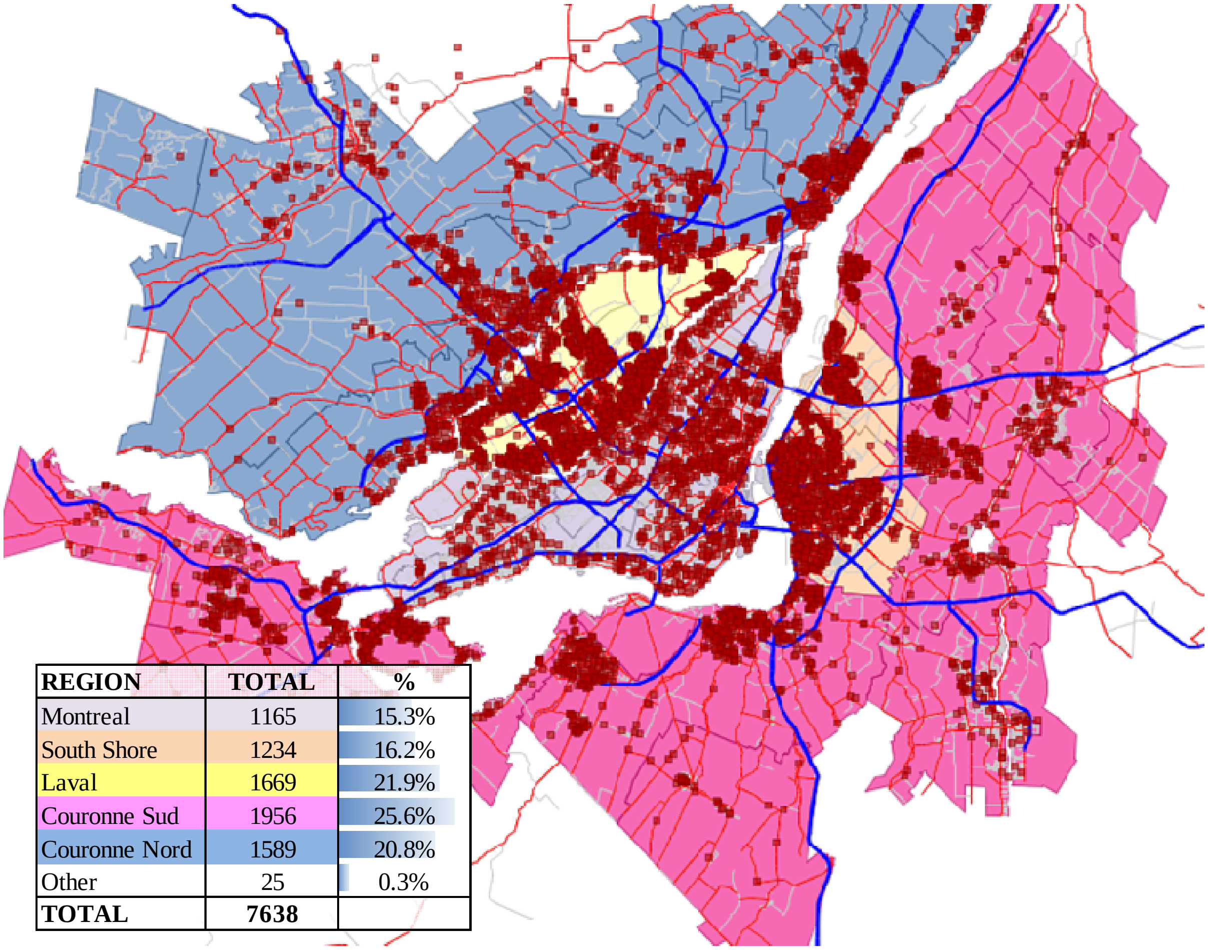


Figure 2.22: Comparison of traffic counts and validated survey responses

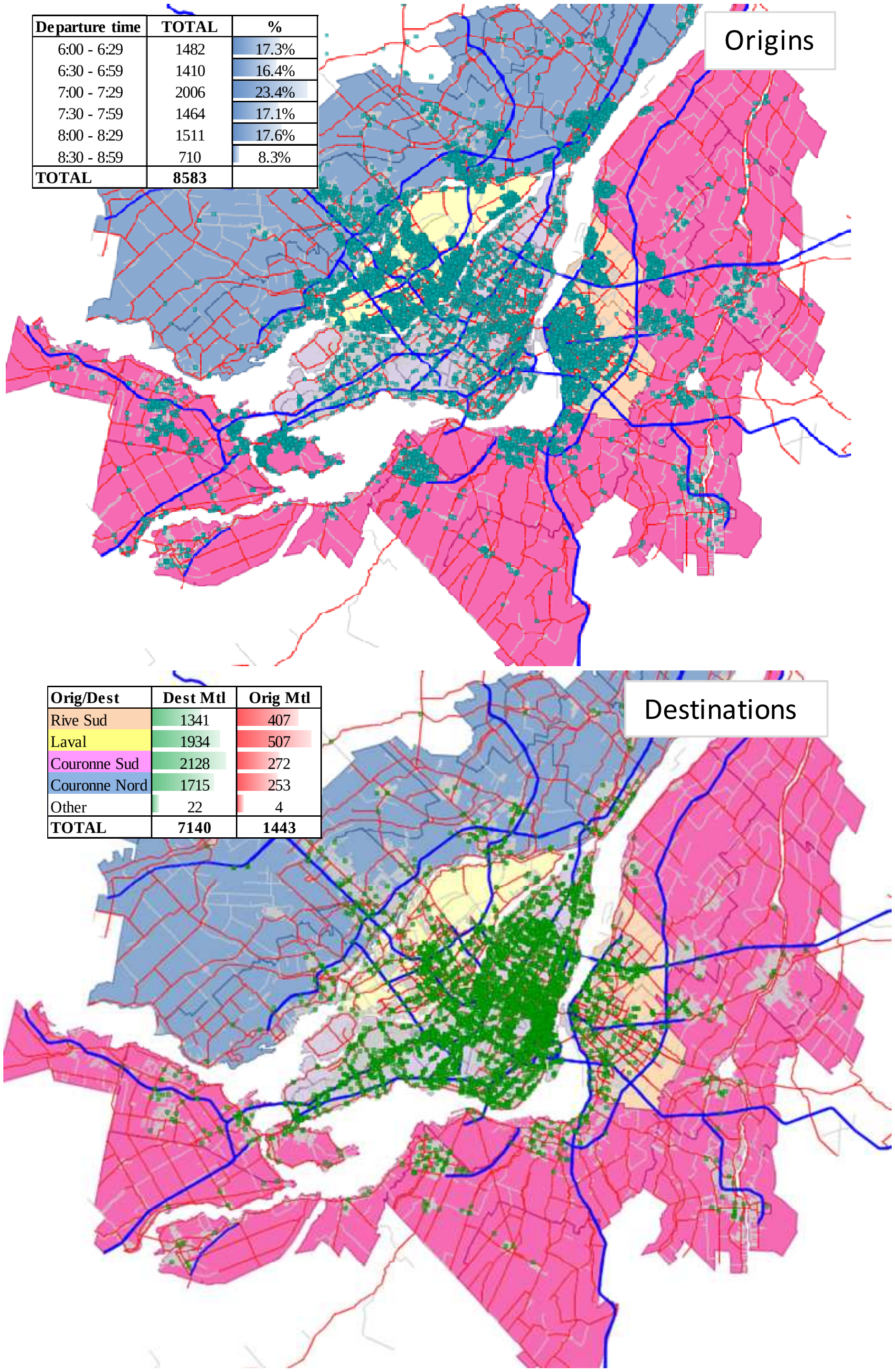
The sample of validated bridge responses in the travel survey corresponds to 7,683 households. Figure 2.23 shows the spatial distribution of these households. The inset table shows their distribution aggregated by region. The island of Montreal is home to 15% of a.m. peak period bridge users. The *couronne sud* accounts for more than a quarter of bridge-user households. The South Shore represents a smaller market (15.6%) than the other suburban regions. Laval and the *couronne nord* together account for two fifths of all bridge-using households.



REGION	TOTAL	%
Montreal	1165	15.3%
South Shore	1234	16.2%
Laval	1669	21.9%
Couronne Sud	1956	25.6%
Couronne Nord	1589	20.8%
Other	25	0.3%
TOTAL	**7638**	

Figure 2.23: Location of households of a.m. peak period bridge users

The spatiotemporal distribution of validated responses is shown in Figure 2.24. Of the 8,583 trips, 7,140 originate off the island of Montreal. Since only single-bridge trips were retained, all these trips terminate somewhere on the island. The remaining 1,443 observations constitute “reverse-commute” trips which originate on the island. The largest supplier of inbound automobile commuters is the *couronne sud* which accounts for 29.8% of all Montreal-bound trips. Laval is the most popular destination for reverse-commuters, attracting 35% of all outbound trips. The distribution of departure times is also noteworthy. Four half-hour intervals each contain around 17% of all departures. The 7:00-7:30 interval captures 23.4% of all departures. After 8:30, demand drops off dramatically. Only 8.3% of travellers declared a departure time within this interval. This finding suggests a constrained arrival time, which would fit with the intuition that most people are expected to be at work at or at school before 9:00.



Departure time	TOTAL	%
6:00 - 6:29	1482	17.3%
6:30 - 6:59	1410	16.4%
7:00 - 7:29	2006	23.4%
7:30 - 7:59	1464	17.1%
8:00 - 8:29	1511	17.6%
8:30 - 8:59	710	8.3%
TOTAL	**8583**	

Orig/Dest	Dest Mtl	Orig Mtl
Rive Sud	1341	407
Laval	1934	507
Couronne Sud	2128	272
Couronne Nord	1715	253
Other	22	4
TOTAL	**7140**	**1443**

Figure 2.24: Origins and destinations of trips with a declared bridge during the a.m. peak period

2.4.4 Factors influencing the route choice of drivers

Once a valid sample of trips has been isolated, an exploratory analysis of the factors which explain the observed behaviour may begin. The behaviour of interest here is the choice of bridge made by auto-drivers and this choice is related to the choice of path. In a detailed urban network composed of many thousands of links, the theoretical number of *possible* acyclic paths connecting a given origin-destination pair is finite but often very large. The size of the path choice set can be reduced by ignoring routes which are much longer than the shortest route. The remaining paths can be considered plausible options. In principle however, for a given origin-destination pair, two paths which differ by a single link and whose travel times are similar constitute two distinct routing options. Therefore, in a detailed network, the size of the plausible path set for a single o-d pair may still be considerable. Much research has focused on the construction of a set of plausible paths and the computation of the choice probability for each. The present research proposes a simpler method which is based on the hypothesis that real drivers do not examine a multitude of alternative routes. The methodology is justified by the ultimate purpose of the simulation model, which is to analyse the usage patterns of major infrastructure elements.

If only major infrastructure elements are considered in the choice model, then the model becomes a representation of facility choice rather than route choice. In the present research, the facility is the major bridge but the approach could be extended to freeways or arterial roads where a representative sample of revealed preference information exists. This approach has two favourable characteristics. First, the set of plausible facilities to be used in the completion of a trip is much more limited than the set of plausible routes. Second, the available alternatives are mostly independent of each other. Overlapping alternatives are rare in the case of freeways and non-existent in the case of bridges. As result, there is no need to account for correlation between alternatives in the latter case.

As a preliminary step in the construction of a predictive model, this section outlines a method for identifying the factors which contribute to the facility choice of drivers. Model development is an iterative process. In the present research, it begins with a simple hypothesis about driver behaviour. The initial model is augmented by examining its incorrect predictions. The framework for comparing model forecasts to observed reality is elaborated in the first subsection.

The second subsection describes a general classification of model errors. The second section demonstrates the utility of the confusion matrix for identifying incorrectly modelled transport markets and analysing them in detail. The third section examines the type of model errors that are most difficult to correct: those that arise from situations of indifference.

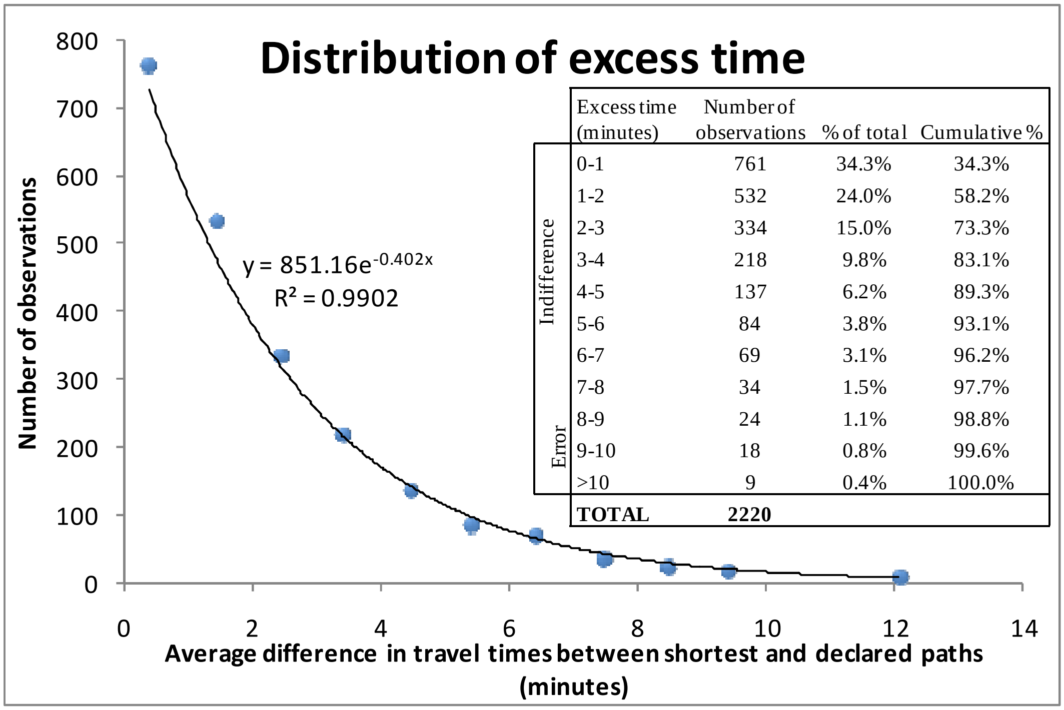
2.4.4.1 The validation and simulation models

In section 2.4.3.3, two simulations were performed for the purpose of isolating a valid sample of bridge declarations. Using the validated sample, the process is repeated. An initial simulation assigns the sampled trips to the network based on the hypothesis that each driver follows the path having the smallest travel time. This unconstrained all-or-nothing assignment is dubbed the simulation model. The second assignment adopts the same hypothesis with respect to driver route choice but constrains the route to the bridge declared in the survey and is called the validation model. Note that if a user-equilibrium does exist, then the validation model is its representation, at least with respect to the major bridges. Evidently, the routes generated by the simulation and validation models are in many cases identical but there will be a significant number of trips whose choice of bridge does not correspond to the simulated shortest path. These trips constitute the errors of the model.

2.4.4.2 Indifference, deviance and error

Sources of error in a model of human behaviour can be segregated into three broad categories: gross errors, deviance, and indifference. Gross errors are cases where human intervention in the observation and model construction process leads to inconsistencies in the data structure. Examples include erroneous declarations by the survey respondent, improper codification of the declared trip by the interviewer and the inaccurate codification of the model network to which the trip is assigned. Deviant behaviour describes cases where the traveller simply did not adopt the behaviour hypothesised by the model (i.e. he did not choose the shortest path). Finally, indifference effects arise when the difference between two alternatives is indistinguishable, both to the driver and to the model. Consider an example where the simulated travel time using bridge 1301 is 13.6 minutes, the simulated travel time by bridge 1302 is 14.1 minutes and the driver chooses bridge 1302. Because the variability of real travel time on a road network and the uncertainty of the estimates generated by the model combine to form a confidence interval which is much larger than 30 seconds, it is impossible to definitively conclude that the driver did not choose the shortest path or that the model made an incorrect prediction.

Since the sample of bridge declarations has already been validated through the elimination of obviously inconsistent observations, it is hoped that the number of incorrect predictions resulting from gross error will be very small. This assertion is tested by plotting the distribution of excess time of incorrectly modelled trips. Excess time is defined here as the difference between the travel time on the route generated by the validation model and the time generated using the simulation model. Figure 2.25 shows the distribution of excess time for the 2,220 travellers who declared a path which was not predicted by the all-or-nothing assignment model. Therefore, all the travellers represented in the figure chose a bridge which, according to the model, generates an itinerary that is longer than the shortest path. The negative exponential form of the distribution (reminiscent of Dial's path choice model (Dial, 1971)) is evident. Only 9 of these trips have excess times greater than 10 minutes and only 10.7% have excess times greater than 5 minutes. Although it is impossible to delineate precisely the boundaries between error, deviance and indifference, it is safe to say that the phenomenon of indifference is much more prevalent than the phenomenon of gross error. This finding would seem to validate most of the partial path information provided by survey respondents. If deviant behaviour is arbitrarily defined by a range of excess time between 2 and 5 minutes, then it accounts for around 31% of the incorrectly modelled trips, which corresponds to 8% of all simulated trips. Indifference, as defined by an excess time of less than 2 minutes, would comprise 58% of erroneous predictions.



	Excess time (minutes)	Number of observations	% of total	Cumulative %
Indifference	0-1	761	34.3%	34.3%
	1-2	532	24.0%	58.2%
	2-3	334	15.0%	73.3%
	3-4	218	9.8%	83.1%
	4-5	137	6.2%	89.3%
	5-6	84	3.8%	93.1%
	6-7	69	3.1%	96.2%
Error	7-8	34	1.5%	97.7%
	8-9	24	1.1%	98.8%
	9-10	18	0.8%	99.6%
	>10	9	0.4%	100.0%
	TOTAL	**2220**		

Figure 2.25: Distribution of travel time differences relative to the shortest path.

2.4.4.3 The Confusion Matrix

Developed in the domain of machine learning (Kohavi & Provost, 1998), the confusion matrix is ideally suited to the evaluation of discrete choice models since it reveals exactly *how* the model makes erroneous predictions. It is sometimes referred to as a prediction-success table in the discrete choice modelling literature. The confusion matrix allows for the calculation of numerous indicators of model performance based on a comparison of observed and predicted outcomes. Table 2.7 is an example of a confusion matrix for evaluating a model that predicts one of two possible results. The row labels represent the observed outcomes and the column labels represent the predicted outcomes. The cell values (*a*, *b*, *c*, and *d*) contain the frequency of each observed-predicted combination. Values along the diagonal (*a* and *b*) contain the correct predictions of the model. The formulae for the calculation of four possible performance indicators are shown below the table.

Table 2.7 : An example confusion matrix

	Predicted		
Observed		1	2
	1	*a*	*b*
	2	*c*	*d*

Correct prediction rate or true positive rate for outcome 1 (T_1):

$$T_1 = \frac{a}{a+b} \tag{2.1}$$

Precision for outcome 1 (P_1):

$$P_1 = \frac{a}{a+c} \tag{2.2}$$

Percentage error (E):

$$E_1 = \frac{(a+c)-(a+b)}{a+b} = \frac{c-b}{a+b} \tag{2.3}$$

Global correct prediction rate or accuracy (A):

$$A = \frac{a+d}{a+b+c+d} \tag{2.4}$$

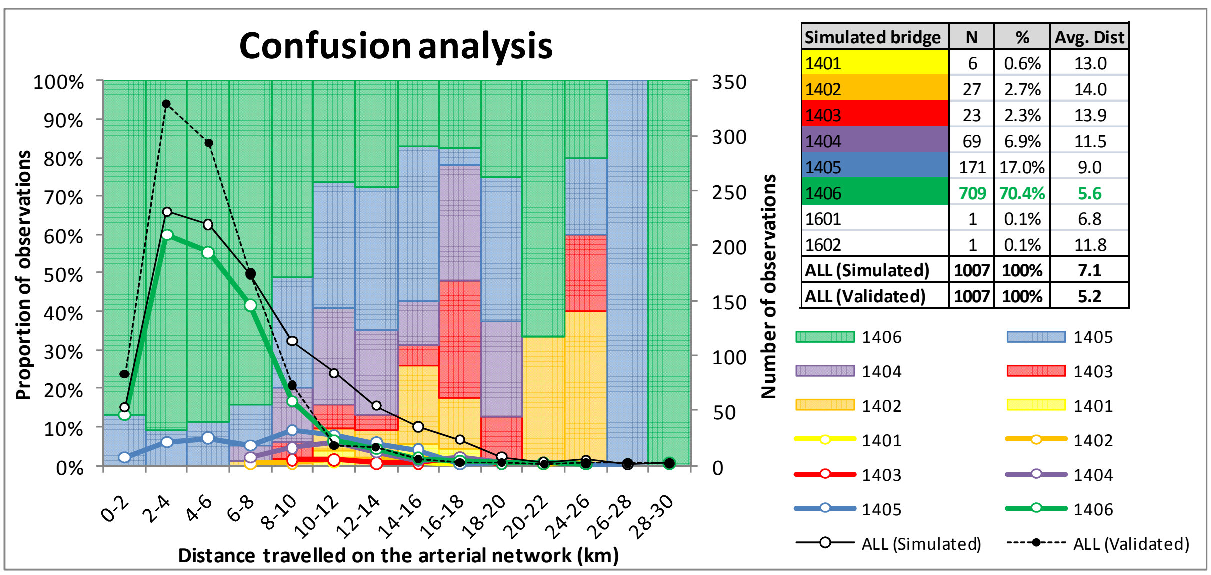
Table 2.8 is a confusion matrix for the 15 major bridges of Montreal. The model being evaluated is the all-or-nothing assignment of trips to the Poly network using the TRANSIMS simulation software (for details, see section 2.5.2). The rows of the matrix represent the observed behaviour, the columns represent the predictions of the model, and the matrix diagonal contains all the correct predictions. So, for example, there were 778 travellers who chose bridge 1301 and 623 travellers who were predicted to use bridge 1301 based on the all-or-nothing assignment. The bridge choices predicted by the model match observed bridge choices in 516 cases. The number of correct predictions divided by the total number of observations is the "correct prediction rate". For bridge 1301, it has a value of 68.8%. The "percentage error" is the predicted bridge volume minus the observed bridge volume all divided by the observed volume. The global correct prediction rate for all 15 bridges is 74.1%. Comparisons of the vector of total observations with the vector of total predictions yield the coefficient of determination (R^2) and the %-RMSE, both standard performance indicators in traffic assignment modelling when comparing observed and simulated traffic volumes. Note that neither of these last two indicators are adequate measures for evaluating a model of choice since they only compare *total* observed and predicted bridge flows.

Table 2.8: Confusion matrix for a.m. peak period trips destined to Montreal Island simulated using an all-or-nothing assignment to an uncongested network.

Number of trips		Modeled bridge																
Screenline	**Observed bridge**	1301	1302	1303	1304	1501	1401	1402	1403	1404	1405	1406	1503	1504	1601	1602	TOTAL	% CORRECT
South Shore	1301	516	85	133	20	24											778	66.3%
	1302	46	186	118	7	3											360	51.7%
	1303	36	67	703	20	2											828	84.9%
	1304	2	2	129	581												714	81.4%
	1501	23		4	4	512							4	2			549	93.3%
Laval	1401						130	53	13	2	38	5			1	1	243	53.5%
	1402						33	289	136		96	2			1	3	560	51.6%
	1403						9	73	514		46	3				9	654	78.6%
	1404						13	11	2	91	77	18				1	213	42.7%
	1405						49	96	68	30	710	64				1	1018	69.7%
	1406						6	27	23	69	171	709			1	1	1007	70.4%
West	1503												384	24			408	94.1%
	1504					1							46	464			511	90.8%
East	1601														37	98	135	27.4%
	1602								56		2	1			9	537	605	88.8%
	TOTAL	623	340	1087	632	542	240	549	812	192	1140	802	434	490	49	651	8583	**74.1%**
	% ERROR	-20%	-6%	31%	-11%	-1%	-1%	-2%	24%	-10%	12%	-20%	6%	-4%	-64%	8%	**85.7%**	**20.4%**
																	R^2	RMSE

The primary interest of the confusion matrix, however, is that it allows for the isolation of particular market segments (Spurr & Chapleau, 2007). All the erroneous predictions of the model are found in the off-diagonal cells and some of these cells contain a significant volume of trips. For example, there were 171 trips which, in reality, used bridge 1406 (the Louis-Bisson Bridge) but were assigned by the model to bridge 1405 (the Médéric-Martin Bridge). A detailed analysis of the trips that make up this improperly modelled market segment may provide insight into the driver decision-making process.

Both the Médéric-Martin and Louis-Bisson bridges carry freeways, belong to the same screenline and run parallel to each other at a distance of approximately 5 km. Driver indifference toward these two alternatives seems likely. A comparison of the attributes of the correctly predicted trips with the attributes of the incorrectly predicted trips allows for the identification of a decision variable not considered by the model. Figure 2.26 is a visualization of this comparison. It shows the distribution of distance travelled on the arterial road network by all the drivers who declared the use of the Louis-Bisson Bridge. A large proportion (70.4%) of these trips was correctly assigned by the simulation model. The remaining trips were distributed among the seven other bridges of the Laval and East screenlines. A separate distribution is plotted for each simulated bridge. The distributions indicate that correctly assigned trips travel shorter distances on the arterial network than the incorrectly assigned trips. This assertion is confirmed by the computation of average distances in the inset table. The distance distribution generated by the validation model is also plotted. It has the same shape as the distribution of trips correctly predicted by the simulation model. These findings imply that drivers do not simply try to minimize their total travel time, but attempt to minimize their use of non-freeway infrastructure. While this result might be arrived at by some other method, the confusion matrix allows for a precise and structured examination of complex behavioural patterns.



Simulated bridge	N	%	Avg. Dist
1401	6	0.6%	13.0
1402	27	2.7%	14.0
1403	23	2.3%	13.9
1404	69	6.9%	11.5
1405	171	17.0%	9.0
1406	709	70.4%	5.6
1601	1	0.1%	6.8
1602	1	0.1%	11.8
ALL (Simulated)	**1007**	**100%**	**7.1**
ALL (Validated)	**1007**	**100%**	**5.2**

Figure 2.26: Detailed analysis of predicted choices of Louis-Bisson Bridge users

2.4.4.4 Detailed analysis of indifference

Situations of indifference are not easily handled by algebraic models. Mathematically, the concept of indifference implies the subtraction of one quantity from another for the purposes of comparison. If the result of the calculation is large in magnitude, then the two quantities are easily distinguished. If the result of the calculation is zero or relatively close to zero, it is impossible for a model to reliably determine which option will be chosen unless additional information is provided. Deterministic methods decide by ignoring the size of the difference between alternatives. Probabilistic methods avoid choosing by assigning probabilities to each of the alternatives. Neither method is an accurate representation of reality. Given these difficulties, the goal of this section is to increase the contrast between competing alternatives in order to reduce the regions of indifference within the space defined by the decision variables.

The geographical expression of indifference between alternatives can be illustrated by the construction of a drainage basin or catchment area for each bridge. Regions where catchment areas meet (decisional watersheds) constitute areas of spatial indifference. The road network is constructed in such a way that any point in the urban region is closest (as measured by distance or by time) to a particular bridge entrance. Similarly, every point is closest to a particular bridge exit. Each trip origin and destination lies, respectively, in an access or egress basin, although the basins of the origin and destination often do not belong to the same bridge. It is not immediately clear which basin determines the choice of bridge. Basins can be constructed using the results of the all-or-nothing shortest path assignment of trips from their origins to the nearest bridge entrance (Figure 2.27). For the purposes of visualization, a grid composed of square kilometre cells was constructed. Each cell of a particular colour contains at least one trip origin within the basin of the corresponding bridge. It is possible for basins to overlap. The size of the basin for each bridge is indicated in square kilometres in the map legend. The figure demonstrates that the size of the catchment area depends on its position relative to other bridges and on the type of facility carried by the bridge. For example, the Lachapelle Bridge has a catchment area of 34 square kilometres and the catchment area of the nearby Médéric-Martin Bridge 155 square kilometres. The difference is because the Médéric-Martin Bridge carries a freeway and the Lachapelle Bridge carries an arterial road. The role of geography is evident in a comparison of the Mercier and Jacques-Cartier Bridges. Neither bridge carries a freeway but the Jacques-Cartier is located within a few kilometres of three other bridges whereas the Mercier is much more isolated. The catchment area of the Mercier Bridge is therefore nearly four times larger than that of the Jacques-Cartier.

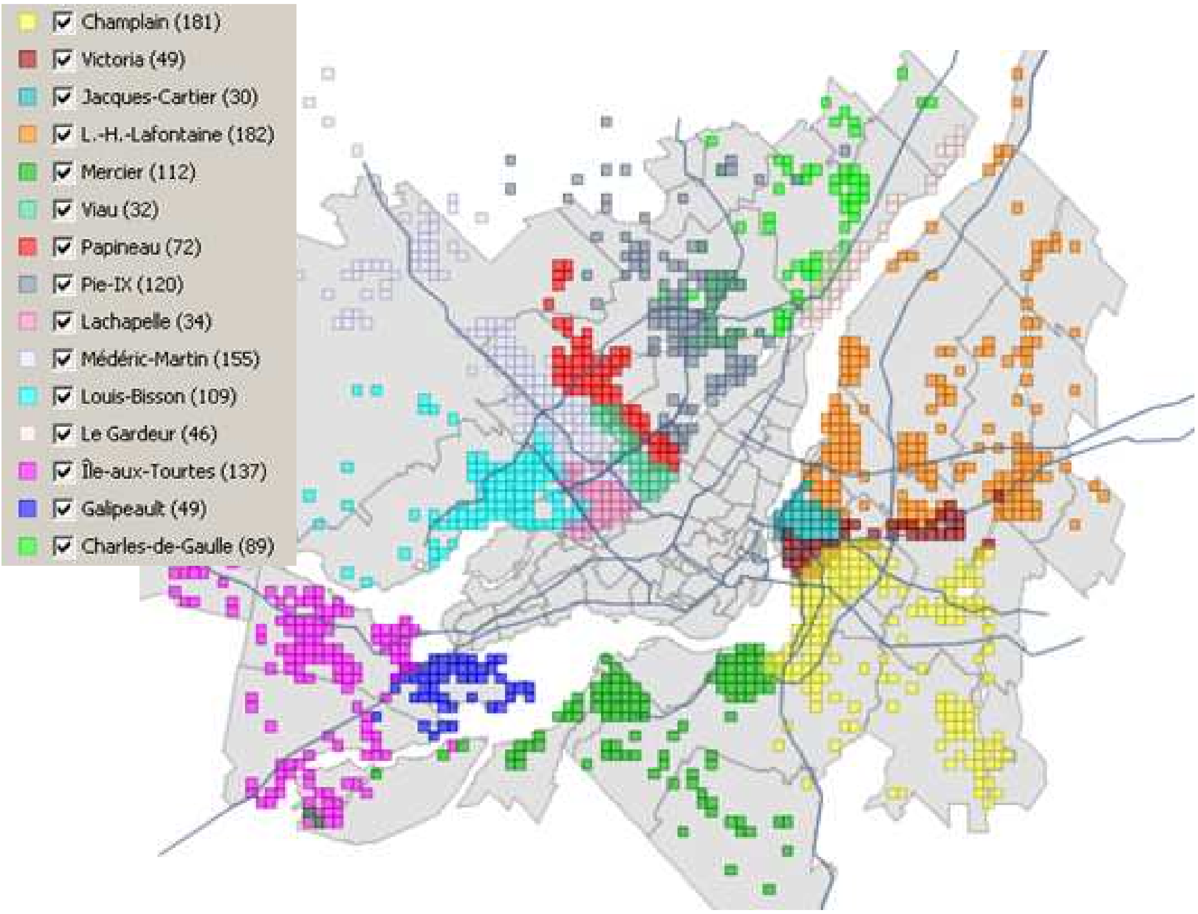


Figure 2.27: Illustration of access catchment areas for all 15 bridges for travel toward Montreal

A simple mathematical formula which quantifies indifference along the dimension of a particular choice variable is:

$$C = x_1 - x_0 \qquad (2.5)$$

Where C is the excess value of the decision variable, x_1 is the value of the decision variable for alternative 1 and x_0 is the value of decision x for alternative 0. In this exercise, alternative 0 will be a presumed optimal (simulated) alternative. A C of small magnitude indicates indifference between the two alternatives. The sign of C indicates which alternative will be chosen, depending on whether the optimal value of x is a maximum or a minimum. Indifference is examined over multiple dimensions by selecting multiple route attributes for the calculation of C. Obvious candidate decision variables are travel time, travel distance and average travel speed.

Figure 2.28 is a graphical demonstration of the formula above. Four possible decision variables are represented simultaneously (excess travel time, excess travel distance, excess speed, departure time). Each point represents an inbound trip assigned by the all-or-nothing assignment to a bridge other than the one that was declared. The pair of alternatives being considered therefore consists of the bridge belonging to the shortest-path itinerary and the bridge declared in the survey. The vertical axis represents the difference between the length of the declared route and the length of the minimal time path (excess distance). The horizontal axis represents difference in travel time between the declared path and the minimal time path (excess time). The colour of each point represents the average excess speed of the trip. Red points indicate trips for which the average speed is much larger than that experienced on the minimum-time path (excess average speed). Blue points are trips for which the average speed is much smaller. Finally, the size of each point represents the time of declared departure in units of minutes after 6 a.m. The expected correlation between excess distance and excess time is evident. In addition, it seems that the variation in speed difference decreases with increasing excess time. When the excess time is large, there is no possibility that the chosen path could be faster than the optimal path. The departure time has no apparent correlation with excess distance, excess time or excess average speed.

In section 2.4.4.2, indifference with respect to travel time was defined as an excess travel time between 0 and 2 minutes. These limits are indicated on the graphic. Another arbitrary definition of indifference is applied to distance: 1 km longer or shorter than the optimal path. Similarly for average speed, the region of indifference is assumed to lie between $\pm$ 9.6 km/h relative to the average speed of the optimal path.

The figure shows that there are 505 trips which fall within a region of generalized indifference where excess speed, time and distance are all close to 0 (region A). But 90 of the travellers who are indifferent to travel time have chosen paths which are significantly shorter, in terms of distance, than the minimum-time path. These trips are located in region B. Also, 31 additional travellers, represented by the red and orange points in region C, choose longer routes travelled at much higher speeds. If the speed indifference threshold is reduced from 9.6 km/h to 6.4 km/h, the number of trips in region C rises to 131. The optimization with respect to speed is not necessarily equivalent to the minimization of travel time since it is possible to find a path which takes longer but also covers a greater distance than the minimum-time path and therefore yields a higher average speed. For example, some drivers may choose the freeway over the arterial network because, even though it takes a few minutes more, it requires less starting and stopping at traffic lights and less interaction with other vehicles. The speed maximization hypothesis could also be stated as the maximization of distance spent on the superior network.

Although the definition of indifference thresholds is arbitrary, this exercise demonstrates that alternative paths which are virtually equivalent along one dimension (such as travel time) are distinct along another dimension (such as speed or distance). Moreover, the analysis of indifference permits the identification of two additional types of optimization behaviour: the minimization of distance and the maximization of average speed.

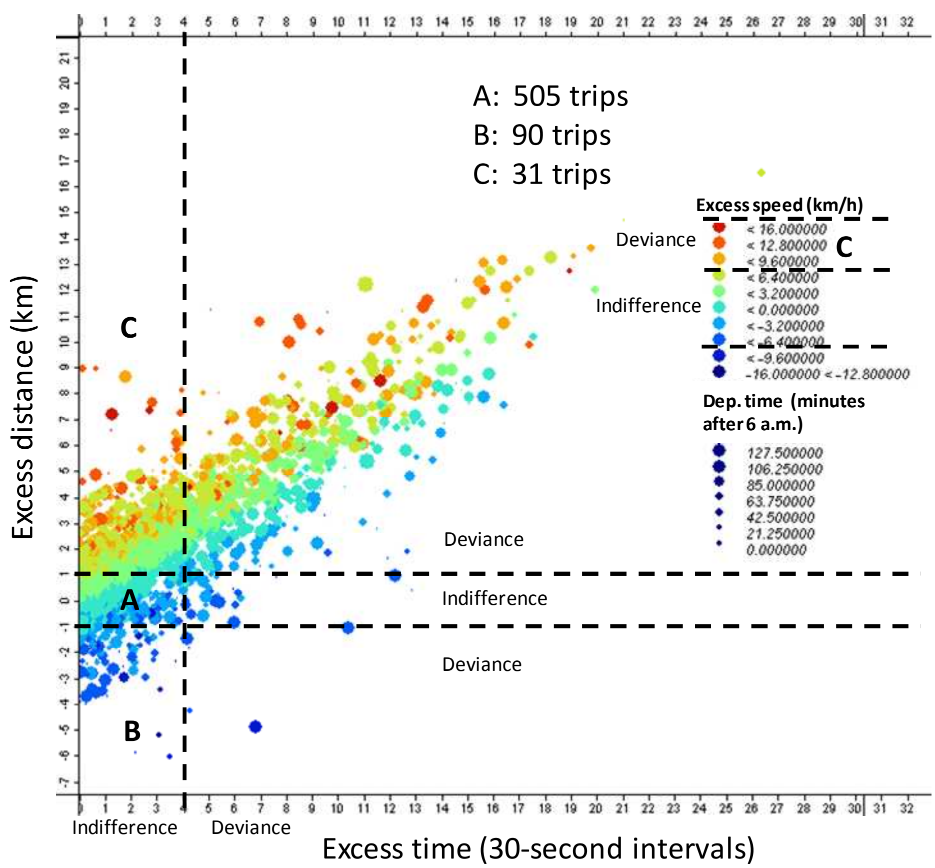


Figure 2.28: Multidimensional analysis of decision variables: differences between the chosen alternative and the shortest-path alternative.

In a situation of indifference, a route might be chosen based on considerations that may have nothing to do with the optimization of cost or utility. These considerations depend upon the psychology and cognitive process of the driver – attributes which are not directly observable in the survey. A comparison of the distribution of personal and cost-independent trip attributes of travellers whose choice of bridge was correctly predicted with the personal attributes of travellers whose choice was not correctly predicted reveals almost no appreciable differences (Table 2.9). The only factor that displays distinct distributions between the two sets of trips is the territory of trip origin. The region of Laval accounts for 31.4% of incorrect predictions and 19.4% of correct predictions. This means that the all-or-nothing assignment model has particular difficulty predicting the bridges used by trips originating in Laval. Conversely, the *Couronne Sud* region contains 28.6% of correctly predicted choices and 13.9% of the erroneous predictions. In other words, the model is especially apt at predicting the choice of bridge for trips originating in the *Couronne Sud*.

Table 2.9: Comparison of the attributes of correctly and incorrectly modelled trips

Attribute	**Correct**	**Incorrect**	**Total cases**
Age group - Men			
15-24	5%	6%	**292**
25-34	20%	21%	**1043**
35-44	33%	33%	**1713**
45-54	28%	28%	**1463**
55-64	12%	11%	**602**
65-74	2%	1%	**82**
75-84	0%	0%	**10**
95-105	0%	0%	**0**
Age group - Women			
15-24	7%	7%	**242**
25-34	23%	26%	**809**
35-44	36%	33%	**1192**
45-54	26%	28%	**899**
55-64	7%	6%	**220**
65-74	0%	0%	**15**
75-84	0%	0%	**0**
95-105	0%	0%	**1**
Professional status			
Full-time worker	89.1%	90.5%	**7679**
Part-time worker	3.5%	2.8%	**284**
Student	3.7%	3.9%	**323**
Retired	1.9%	1.3%	**151**
Other	1.4%	1.3%	**115**
At home	0.4%	0.3%	**31**
Departure time			
6:00-6:29	17.7%	16.0%	**1482**
6:30-6:59	16.4%	16.5%	**1410**
7:00-7:29	23.0%	24.3%	**2006**
7:30-7:59	16.9%	17.5%	**1464**
8:00-8:29	17.6%	17.6%	**1511**
8:30-8:59	8.4%	8.0%	**710**
Trip Purpose			
Work	84.0%	85.0%	**7235**
Study	3.5%	3.4%	**298**
Return home	2.9%	1.8%	**222**
Leisure	1.2%	1.0%	**100**
Shopping	0.3%	0.1%	**21**
Other	8.1%	8.6%	**707**
Region of origin			
Montreal	16.6%	17.4%	**1443**
South Shore	15.6%	15.6%	**1341**
Laval	19.4%	31.4%	**1934**
Couronne sud	28.6%	13.9%	**2128**
Couronne nord	19.5%	21.4%	**1715**
Total cases	**6363**	**2220**	**8583**

The reasons for these aberrations relate to the characteristics of the bridges used in each case, as shown in Table 2.10. Trips originating in Laval represent the largest group of drivers using non-freeway bridges (348 declarations). Trips originating in the *Couronne sud*, meanwhile, represent the largest group of drivers using freeway-carrying bridges (2,032 declarations). As is made apparent by its confusion matrix (Table 2.8), the all-or-nothing assignment model has difficulty accurately reproducing the use of non-freeway bridges, especially 1401, 1404 and 1601. The first two are used extensively by drivers who begin their trip in Laval. Freeway-carrying bridges are generally better represented in the model, and are especially well-represented in the West screenline (bridges 1503 and 1504) which primarily serves trips originating in the *Couronne sud*. The %-error statistics in the same confusion matrix show that, with the exception of bridge 1601, the predicted volumes on non-freeway bridges are not dramatically different from the observed volumes. The total predicted volumes are fairly good, but an important proportion of these volumes consist of trips assigned to an incorrect bridge.

Table 2.10: Observed use of freeway and non-freeway bridges by screenline

	Declared bridges	
Region of origin	**Freeway**	**Non-freeway**
Montreal	1363	80
South Shore	1077	264
Laval	1586	348
Couronne sud	2032	96
Couronne nord	1574	163
ALL	**7632**	**951**

2.4.4.5 Indifference and captivity

So far, the analysis of indifference has mostly involved the comparison between the preferred alternative as identified by a simulation model with the chosen alternative observed in the survey. The more conventional approach involves comparisons between observed alternatives. Such comparisons form the basis of most mathematical models of choice. Figure 2.29 is an illustration of observed driver behaviour when the two alternatives are the Champlain and Jacques-Cartier bridges. The graphic shows market share of each bridge as a function of the ratio of travel-times. To construct this ratio, each of the 1326 trips observed on either the Jacques-Cartier or the Champlain were assigned to the network twice. In the first assignment, the trips are assigned to the bridge declared in the survey. In the second assignment they are assigned to the other bridge. For a given trip, the travel time ratio is the simulated time required to complete the trip using the Jacques-Cartier divided by the simulated time resulting from the use of the Champlain. The trips are grouped into bins of 0.05 to construct the graphic.

A region of indifference - where the travel time ratio is close to 1 the market share of each bridge is close to 50% - is visible in the figure. There are also two regions of captivity where the market share of each bridge is 100% and the travel time ratio is very different from 1. Despite the presence of some deviant points, the graphic lends additional credibility to the bridge declarations in the survey. The shape of the curves is highly reminiscent of the logistic distribution which suggests the adoption of a probabilistic approach. Such an approach is described in the next section.

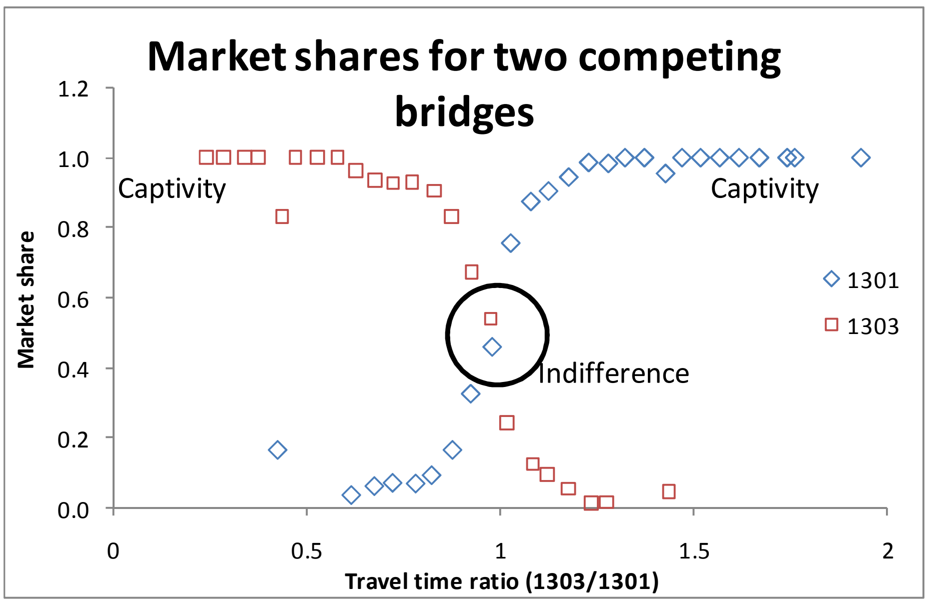


Figure 2.29: Bridge market share as a function of travel time

2.5 Disaggregate Simulation Models

Once declared responses have been validated, the information they contain can be used to construct a coherent model of bridge choice. Following common practice, the model is based on three primary elements: the representation of demand, the representation of supply and a mechanism for relating the two. One such mechanism has already been presented in the form of an all-or-nothing assignment of trips to a detailed and completely uncongested network. This simple model served primarily to test the validity of the travel survey responses and to illustrate how multiple considerations may play a role in the determination of facility choice. With a correct prediction rate of over 74%, the model performs well despite its simplicity. The purpose of this section is to estimate models which consider additional elements believed to play a role in the choice of route on a road network. These elements include traffic congestion, control systems and the hierarchical structure of the network.

Conventional methods of assigning travel demand to networks are aggregate. The units of analysis are packets of flow. If the results of any model are to be incorporated in an analysis of infrastructure usage patterns, it is essential that the disaggregate structure of the demand information be retained. Two types of disaggregate models are examined. The first one is the discrete choice random utility approach commonly used in investigations of mode choice and route choice. The second method employs the TRANSIMS software which was designed for the activity-based modelling of complete urban transportation systems.

2.5.1 Random-utility models of facility choice

A discrete choice model is an appropriate method for performing a disaggregate assignment of vehicles to a network since it conserves individuals and their attributes during the route selection process. The model can use either traveller attributes or service attributes as decision variables. In addition, the data validation process suggested that the probability of choosing a particular bridge has a logistic distribution.

Typical applications of discrete choice methods to problems of route choice are significantly more complicated than the model presented here. The reason is that most route choice models consider a large number of alternative paths which are not necessarily independent of each other. Indeed, overlapping routes are extremely common. The large number of alternatives and the correlation between alternatives requires a “path-size” variable to account for this fact. In addition, the model can only be estimated if a set of alternatives is specified beforehand. In the vast majority of cases, the alternative routes are never observed and must be synthesized using, for example, a random-walk algorithm. The number of synthesized alternative paths in the choice set of a single traveller is usually on the order of tens but can be on the order of hundreds. This path-enumeration approach seems behaviourally unrealistic since it is highly unlikely that real travellers are aware of more than three or four alternative paths for a given route. The present model can be simplified because the choice is not between alternative routes but between alternative facilities (bridges). The choice set for each trip need not be synthesized but is instead uniquely determined by the regional geography. The fact that the major bridges form a complete cordon composed of multiple screenlines means that, for any one trip, the number of facilities in the choice set can be no greater than 8 and each facility is completely independent of all the others since only one facility can be chosen per route.

2.5.1.1 Specification of three multinomial logit models

The multinomial logit model is founded on the hypothesis that the decision-makers are utility maximizers and that the utility of each decision-maker has observable and unobservable (random) components. The difference between the observable utility and the real utility constitutes a random-error term which has an assumed expected value of zero. If the distribution of the error term is assumed to be identically and independently Gumbel then the model will take the form of a multinomial logit. In the simplest case, the utility of a decision-maker is a linear function of the choice attributes and/or of the decision-maker himself. In order to estimate utility coefficients for attributes of the alternatives, it is necessary to compute attribute values for all alternatives in the choice set, even though, for a given traveller, only one alternative is observed. In the models presented below, therefore, each traveller is simulated n times, where n is the number of alternative bridges in the traveller's choice set. During each simulation, the traveller is forced onto one of the j bridges forming the screenline which must be crossed to complete the trip.

Discrete choice models are characterized by a binary dependent variable representing the observed choice. This variable (y_{ij}) is 1 if person i chose alternative j and 0 otherwise. The logit model is a type of discrete choice model with the following form (a more complete derivation can be found in (Train, 2003)):

$$\Pr(m_i = j) = \frac{e^{U_{ij}}}{\sum_{l=1}^{k} e^{U_{il}}} \qquad (2.6)$$

where m_i is the alternative chosen by person i, U_{ij} is the utility person i obtains from alternative j and k is the number of alternatives among which person i can choose. Each utility function is a linear combination of explanatory variables ($\mathbf{X}$). The utility weights associated with these variables are the estimated parameters of the model ($\widehat{\boldsymbol{\beta}}$). The estimated utility function of person *i* for alternative *j* ($\widehat{U}_{ij}$) can therefore be expressed as:

$$\widehat{U}_{ij} = \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}} \qquad (2.7)$$

The parameter estimates of the logit model cannot be found using algebra. A maximum likelihood simulation process is adopted to solve for these parameters. The simulation requires the computation of a likelihood function (L) which is interpreted as the joint probability of observing the events that were in fact observed. The likelihood function can therefore be written as the multiplication of the estimated probabilities of each observation.

$$L = \prod_{i=1}^{N}\prod_{l=1}^{k} \mathrm{Pr}(m_i = l)^{y_{il}} \qquad (2.8)$$

To simplify the calculation process, the simulation process is applied to the logarithm of the likelihood function. The function to be maximized becomes a sum rather than a product and is known as the log-likelihood function (LL).

$$LL = \sum_{i=1}^{N}\sum_{l=1}^{k} y_{il} \ln[\mathrm{Pr}(m_i = l)] \qquad (2.9)$$

It so happens that the log-likelihood function is convex and its maximum is therefore found at the point where its first derivative with respect to $\widehat{\boldsymbol{\beta}}$ is equal to 0. This condition is expressed as:

$$\sum_{i=1}^{N}\sum_{l=1}^{k} x_{il}(y_{il} - \mathrm{Pr}(m_i = l)) = 0 \qquad (2.10)$$

or in matrix form :

$$\sum_{i=1}^{N} \mathbf{X}_i \left(\mathbf{Y}_i - \widehat{\mathbf{P}}_i \right) = 0 \qquad (2.11)$$

where $\widehat{\mathbf{P}}_i$ is the vector of estimated choice probabilities for person i. The parameters $\widehat{\boldsymbol{\beta}}$ which satisfy this condition are the maximum-likelihood estimators of the logit model. The statistical significance of these estimators can be evaluated in much the same way as the parameters of linear regression model.

The “pseudo r-squared” is a commonly adopted goodness-of-fit measure for logit models and is based on a comparison of the value of the log-likelihood function using the initial parameter estimates (LL_0) and its maximum value (LL_f). The statistic (ρ^2) always takes a value between 0 and 1 and is computed as

$$\rho^2 = \frac{LL_f - LL_0}{LL_0} \qquad (2.12)$$

One of the interesting properties of the logit model is the possibility of specifying different utility functions for different groups of travellers. It might be suggested, for example, that men and women assign different utilities to travel time. To test this hypothesis, separate travel time parameters could be estimated for men and women. The adoption of this model structure is sometimes referred to as “stratification”.

A large number of multinomial bridge choice models were tested. Only three of the more interesting ones are presented here. The first model estimates separate utility functions for each direction of travel (inbound vs. outbound). The explanatory variables are: the Euclidean distance from the trip origin to the bridge entrance in kilometres (access distance); the Euclidean distance from the bridge exit to the destination in kilometres (egress distance); a variable which is 1 if the bridge is connected to the freeway network and 0 otherwise (freeway network); the free flow travel time of the trip based on an all-or-nothing assignment (free flow travel time); the deviation from the shortest path. This last variable (similar to one constructed by Frejinger (2008)) has a value of 1 if the path in question has the smallest travel time among the available options. If the path in question is not the shortest, then the variable (d) takes a positive value less than 1. It is calculated for bridge j and traveller i as follows:

$$d_{ij} = \frac{\min\,(t \in T_i)}{t_j} \tag{2.13}$$

where t_j is the travel time using bridge j and T_i is the set of travel times corresponding to the bridges available to traveller i. These variables are stratified by direction (inbound or outbound) based on the assumption that drivers leaving Montreal during the morning peak period follow a different decision process than the drivers entering.

The estimation of Model 1 generates a set of probabilities associated with each bridge choice. For the purpose of generating vehicle volumes on the bridges, each vehicle is assigned to the bridge with the highest predicted probability (i.e. the highest estimated utility). This all-or-nothing assignment is used to compute bridge volumes. These predicted volumes are then divided by the observed volumes to generate a "utilization ratio". This new variable is added to the utility functions of Model 2. The dependent variable in Model 2 is the bridge choice pattern predicted by Model 1.

The estimated parameters of both models are shown in Table 2.11. All the parameters are statistically significant at the 5% confidence interval and have the correct sign. The relative magnitudes of the directional access and egress parameters suggest that distance accumulated on the island of Montreal offers considerably more disutility than the off-island portion of the trip. In addition, a bridge connected to the freeway network appears to be more attractive than a bridge which is not. This effect may be related to the greater capacity of freeways relative to arterial roads. The addition of a variable to account for capacity constraints (the utilization ratio) is statistically significant but adds little to the explanatory power of Model 2. The ρ^2 coefficients for both models are close to 60%.

Table 2.11: Estimated parameters of two multinomial logit models of bridge choice

		Model 1			Model 2		
	Value at mean	**Coefficient**	**z**	**p**	**Coefficient**	**z**	**p**
Access distance (km)							
Outbound	7.668	-0.338	-9.36	0.000	-0.399	-10.35	0.000
Inbound	9.677	-0.192	-11.67	0.000	-0.178	-10.48	0.000
Egress distance (km)							
Outbound	8.653	-0.114	-3.08	0.002	-0.137	-3.51	0.000
Inbound	8.210	-0.259	-16.29	0.000	-0.28	-16.92	0.000
Freeway network (1 or 0)							
Outbound	0.620	1.851	18.76	0.000	2.84	23.97	0.000
Inbound	0.581	1.135	28.1	0.000	2.138	28.72	0.000
Freeflow travel time (minutes)							
Outbound	20.411	-0.256	-4.06	0.000	-0.329	-4.94	0.000
Inbound	21.738	-0.159	-5.43	0.000	-0.23	-7.54	0.000
Deviation from shortest path							
Outbound	0.969	4.229	3.26	0.001	3.065	2.29	0.022
Inbound	0.975	6.189	9.44	0.000	5.509	8.21	0.000
Utilization ratio	1.000				-2.196	-16.81	0.000
LL_0		**-13860**			**-13860**		
LL_f		**-5675**			**-5529**		
ρ^2		**59.10%**			**60.10%**		

The relative importance of an explanatory variable can be further evaluated by computing the utility weight at its mean value. A graphic of such an evaluation applied to Model 1 is shown in Figure 2.30. Note that the figure shows absolute weights. No distinction is made between positive and negative values of utility. Of particular interest are the differences between the inbound and outbound utility functions. The most important variable for inbound trips is the deviation from the shortest path which accounts for over 40% of the total utility. For outbound trips, the free-flow travel time dominates. The access and egress variables are also worth noting. A typical outbound driver receives more than twice as much disutility from the access segment than the egress segments. An inbound driver meanwhile receives more disutility from the egress segment than from the access segment. Both the outbound access and inbound egress segments represent portions of the trip which are made on the Island of Montreal. This suggests that the Montreal network, as opposed to the off-island suburban network, plays a dominant role in the choice of bridge. It also suggests that auto-travel on the island is more troublesome than auto-travel off the island.

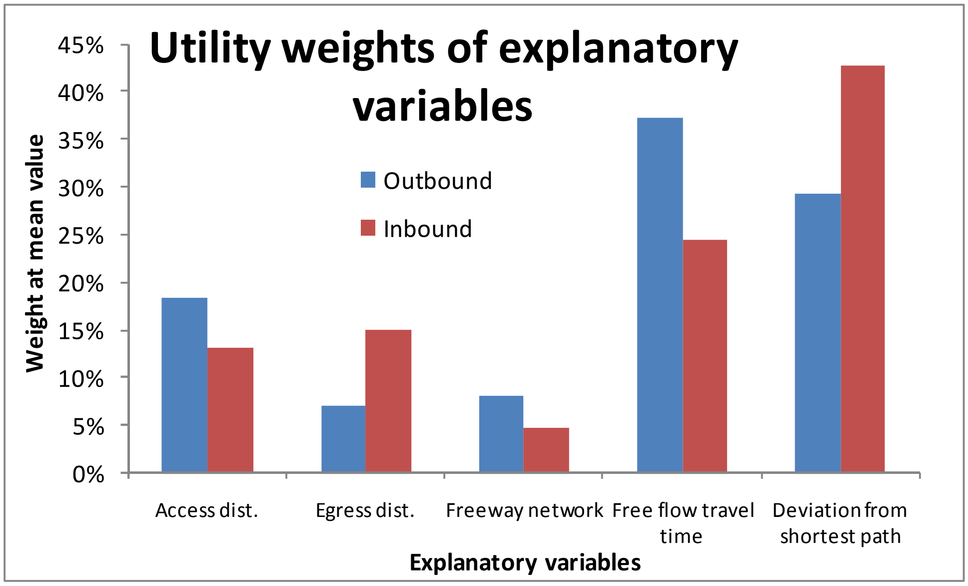


Figure 2.30: Utility weights of explanatory variables for Model 1

A more detailed assessment is provided by the confusion matrix (Table 2.12 and Table 2.13). Note that, by the conventional standard of the correlation coefficient (90% and 96% for Model 1 and Model 2, respectively), both models are excellent for reproducing observed volumes. Considering their simplicity, they also display strong predictive power in that they each reproduce around 75% of observed choices. The effect of the utilization ratio (a proxy indicator of congestion) is clearly seen in the reduction in % error of bridges which do not carry freeways, particularly the Victoria (1302), the Viau (1401) and the Lachapelle (1404). In all three cases, the addition of the utilization ratio brings the % error closer to zero and increases the % correctly predicted. An interesting effect is observed on two hybrid bridges (1402 and 1403, which are classified as freeways at one end and as arterial roads at the other) whose % error becomes zero in Model 2. Despite a better match with observed volumes, however, the correct prediction rate increases only slightly for bridge 1402 and actually decreases in the case of bridge 1403. This finding recalls the more general result that an improvement in the representation of one facility usually comes at the expense of the representation of another. The Jacques-Cartier Bridge (1303) and the Victoria Bridge provide an example of this phenomenon. The correct prediction rate of the Jacques-Cartier falls from 69.9% in Model 1 to 62.8% in Model 2 while the correct prediction rate for the Victoria rises from 39.4% to 45.3%.

Table 2.12: Confusion matrix for Model 1.

Number of trips		**Modeled bridge**																
Screenline	**Observed bridge**	1301	1302	1303	1304	1501	1401	1402	1403	1404	1405	1406	1503	1504	1601	1602	TOTAL	% CORRECT
South Shore	1301	667	25	54	24	8											778	85.7%
	1302	100	142	104	14												360	39.4%
	1303	128	48	579	73												828	69.9%
	1304	15	2	78	619												714	86.7%
	1501	47		2	5	495											549	90.2%
Laval	1401						118	38	15	1	66	5					243	48.6%
	1402						29	285	120		125	1					560	50.9%
	1403						8	71	512		60	3					654	78.3%
	1404						3	7	2	57	123	21					213	26.8%
	1405						36	90	49	4	779	60					1018	76.5%
	1406						1	7	11	20	213	755					1007	75.0%
West	1503												386	22			408	94.6%
	1504												45	466			511	91.2%
East	1601														17	118	135	12.6%
	1602							1	46			2			2	554	605	91.6%
	TOTAL	957	217	817	735	503	195	499	755	82	1366	847	431	488	19	672	8583	**74.9%**
	% ERROR	23%	-40%	-1%	3%	-8%	-20%	-11%	15%	-62%	34%	-16%	6%	-5%	-86%	11%	**90.2%**	**23.6%**
																	R2	RMSE

Table 2.13 : Confusion matrix for Model 2

Number of trips		**Modeled bridge**																
Screenline	**Observed bridge**	1301	1302	1303	1304	1501	1401	1402	1403	1404	1405	1406	1503	1504	1601	1602	TOTAL	% CORRECT
South Shore	1301	668	34	41	27	8											778	85.9%
	1302	97	163	75	25												360	45.3%
	1303	132	74	520	102												828	62.8%
	1304	16	3	61	634												714	88.8%
	1501	51		1	6	491											549	89.4%
Laval	1401						127	34	13	2	60	7					243	52.3%
	1402						36	318	80	1	121	4					560	56.8%
	1403						13	97	484		54	6					654	74.0%
	1404						3	8		72	102	28					213	33.8%
	1405						47	93	31	16	755	76					1018	74.2%
	1406						2	5	9	23	173	795					1007	78.9%
West	1503												386	22			408	94.6%
	1504												43	468			511	91.6%
East	1601														20	115	135	14.8%
	1602						1	3	34		1	2			4	560	605	92.6%
	TOTAL	964	274	698	794	499	229	558	651	114	1266	918	429	490	24	675	8583	**75.3%**
	% ERROR	24%	-24%	-16%	11%	-9%	-6%	0%	0%	-46%	24%	-9%	5%	-4%	-82%	12%	**95.9%**	**19.0%**
																	R2	RMSE

Models 1 and 2 were based on "simple" variables such as straight line distance and free-flow travel time. A more sophisticated formulation incorporates some measure of traffic congestion. In Model 3, the link travel times are based on the data collected using GPS-equipped para-transit vehicles (see section 2.3.3.3.2). Unlike a regular traffic assignment model where network congestion is a predicted output, congestion in this model is an exogenously-specified input based on observation. The travel time of each path is segmented into access, bridge and egress components. In addition to being stratified by direction, the travel time variables are also stratified by screenline. The estimated parameters of Model 3 are shown in Table 2.14. The signs of all travel time variables are negative except where the coefficient is not statistically different from zero (inbound on the West screenline for the bridge portion of the trip). Only two coefficients are not statistically different from 0 at the 95% confidence interval. The pseudo r-squared is just under 61%.

Table 2.14: Estimated parameters of Model 3

Explanatory variable	Direction	Screenline	Value at mean	Coefficient	z	p
Congested access time (minutes)						
	Inbound	S. Shore	17.71	-0.257	-13.56	0.000
		Laval	17.34	-0.342	-19.71	0.000
		West	20.22	-0.127	-1.48	0.139
		East	24.49	-0.126	-3.88	0.000
	Outbound	S. Shore	25.08	-0.210	-30.98	0.000
		Laval	21.26	-0.292	-37.28	0.000
		West	19.89	-0.326	-14.61	0.000
		East	17.23	-0.324	-11.96	0.000
Congested time on the bridge (minutes)						
	Inbound	S. Shore	17.37	-0.166	-2.23	0.026
		Laval	4.09	-0.226	-5.06	0.000
		West	3.77	0.824	1.09	0.277
		East	4.08	-0.204	-3.32	0.001
	Outbound	S. Shore	16.28	-0.124	-11.93	0.000
		Laval	4.42	-0.175	-9.69	0.000
		West	4.49	-0.153	-2.44	0.015
		East	5.96	-0.205	-6.19	0.000
Congested egress time (minutes)						
	Inbound	S. Shore	20.48	-0.172	-12.55	0.000
		Laval	15.99	-0.243	-14.99	0.000
		West	19.69	-0.279	-4.36	0.000
		East	17.69	-0.225	-5.87	0.000
	Outbound	S. Shore	14.83	-0.266	-31.07	0.000
		Laval	18.84	-0.280	-38.61	0.000
		West	25.09	-0.177	-4.72	0.000
		East	32.33	-0.116	-5.99	0.000
LL_0	-13860					
LL_f	-5420					
ρ^2	60.90%					

Figure 2.31 shows the composition of each of the 8 utility functions specified in Model 3. The weight of each explanatory variable is computed using its mean value. The total disutility for each screenline-direction is also shown. For inbound trips, the access portion carries the greatest utility weight for all screenlines. The utility weights for outbound trips display greater variability but the access segments dominate on the South Shore and Laval screenlines. The prevailing importance of the access portions of trips is indicative of the queues which develop upstream of the bridges during the peak period. The total disutility is representative of the inconvenience associated with each screenline. The inbound directions all have higher disutilities than their outbound counterparts due to the higher levels of traffic congestion.

The finding that the access portions of inbound trips carry greater weight in the choice of bridge than the egress portions appears to contradict Models 1 and 2. This inconsistency results from the independence of Euclidean distance from congested travel time. It is possible that the disutility associated with travel on the Island of Montreal in Models 1 and 2 is not related to traffic congestion. The density of traffic signals, the prevalence of one-way streets and turning restrictions as well as the presence of buses and other heavy vehicles may all have an influence on the choice of bridge and their aggregate effect can be captured using the straight-line length of Montreal part of the trip. An average congested travel time, on the other hand, does not necessarily account for these phenomena.

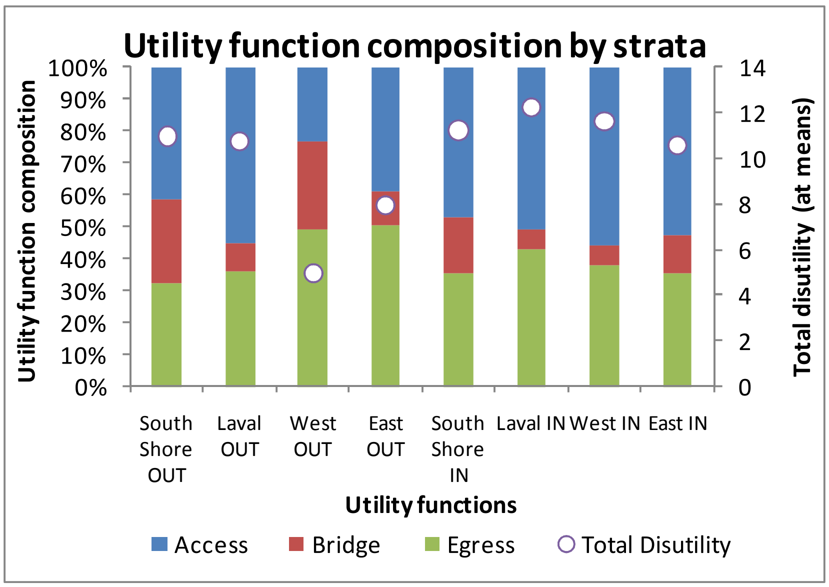


Figure 2.31: Utility function composition for Model 3

The confusion matrix of Model 3 (Table 2.15) demonstrates its predictive power. Nearly 76% of all observed responses were successfully reproduced. With exception of the Victoria Bridge in Model 2, the non-freeway bridges are better represented than in the first two models. With respect to the predicted and observed bridge flows, the coefficient of determination (r-squared) is 98.0%. The small magnitude of these differences is perhaps surprising given that Models 1 and 2 use unrealistic travel times and straight-line distances while Model 3 uses travel times derived from direct observation.

Table 2.15 : Confusion matrix for Model 3

Number of trips		Modeled bridge																
Screenline	**Observed bridge**	1301	1302	1303	1304	1501	1401	1402	1403	1404	1405	1406	1503	1504	1601	1602	TOTAL	% CORRECT
South Shore	1301	610	36	85	24	23											778	78.4%
	1302	85	155	107	9	4											360	43.1%
	1303	85	53	637	51	2											828	76.9%
	1304	7	2	82	619	4											714	86.7%
	1501	25		2	2	520											549	94.7%
Laval	1401						144	39	8	2	41	9					243	59.3%
	1402						58	344	54		89	15					560	61.4%
	1403						19	120	439	1	66	9					654	67.1%
	1404						11	11	1	79	75	36					213	37.1%
	1405						77	107	18	18	664	134					1018	65.2%
	1406						9	15	8	27	101	847					1007	84.1%
West	1503												369	39			408	90.4%
	1504												27	484			511	94.7%
East	1601														23	112	135	17.0%
	1602							1	23		3				14	564	605	93.2%
	TOTAL	812	246	913	705	553	318	637	551	127	1039	1050	396	523	37	676	8583	**75.7%**
	% ERROR	4%	-32%	10%	-1%	1%	31%	14%	-16%	-40%	2%	4%	-3%	2%	-73%	12%	**98.0%**	**12.2%**
																	R2	RMSE

2.5.1.2 Summary

In terms of predictive power, the adoption of a random utility approach to the modelling of bridge choice produced somewhat better results than an all-or-nothing assignment. The hierarchical structure of the road network was found to have a statistically significant influence on the choice process. A proxy indicator of congestion, scaled to account for the sampling rate of the travel survey, was also statistically significant although its explanatory power was small when other variables are included in the model. Free-flow travel time and straight-line distance, two very easily-calculated quantities, were found to have strong explanatory power. These results raise questions about the importance of congestion in the choice of bridge, especially those which carry freeways. The users of these bridges are often captive since the available alternatives, which imply either lengthy detours or greater impedance, are not adequate substitutes even under congested conditions. This realization in turn suggests that the assumption of a user-equilibrium, as traditionally conceived, may be worth re-examining.

Even though the models presented in this section are models of discrete choice, their output can be processed to give results which are identical to those of a typical traffic assignment model. The predicted paths can be aggregated over links to generate link flows (f_a).

$$f_a = \sum_{\Omega} \sum_{k} f_{\Omega,k} \delta^a_{\Omega,k} \text{ where } \delta^a_{\Omega,k} = \begin{cases} 1 \text{ if link } a \text{ is part of path } \Omega,k \\ 0 \text{ otherwise} \end{cases} \tag{2.14}$$

This information can be expressed as a thematic map (Figure 2.32) which is the typical output of traffic assignment simulators.

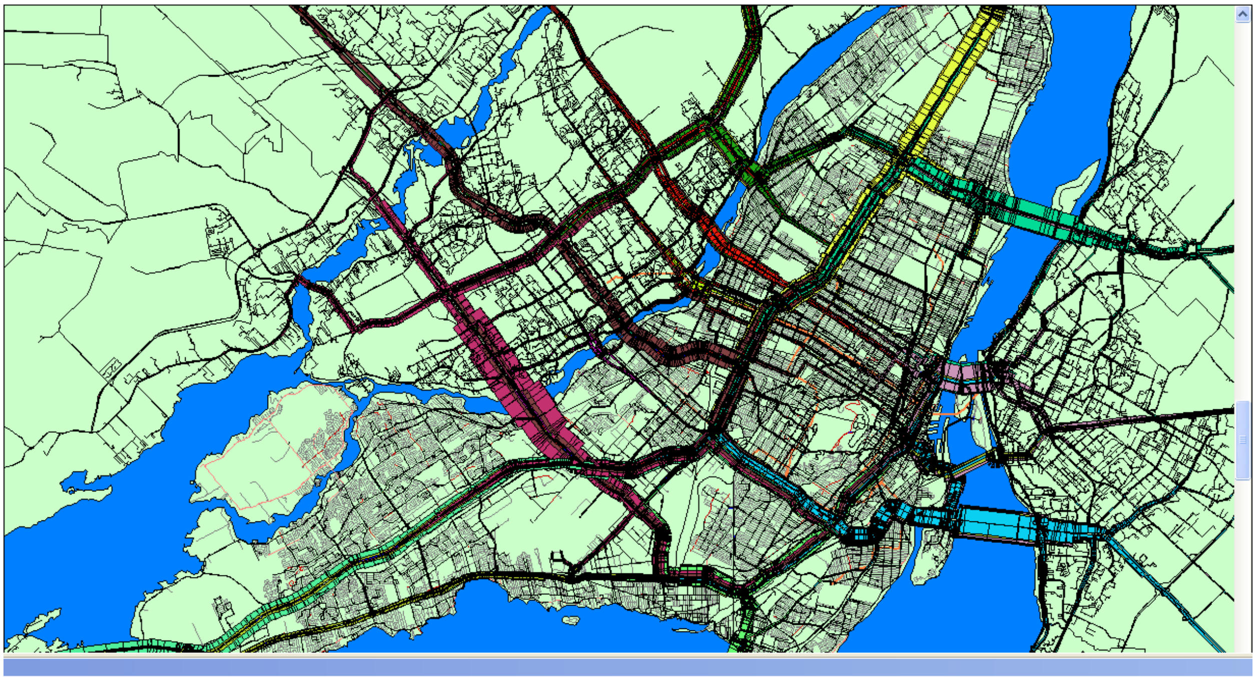


Figure 2.32: The assignment of modelled choices to the detailed network.

The estimated parameters of these discrete choice models provide clues as to the way in which drivers perceive the network. The process just described demonstrates how observed behaviour can suggest unobserved characteristics of the road network, in this case the higher impedance of the Montreal network relative to the suburban network and the existence of queues upstream of some bridges. These phenomena are detectable because the all-or-nothing assignment algorithm generates complete trip itineraries given the choice of bridge predicted by the discrete choice model. Each itinerary has numerous attributes such as the time of departure, the duration, the length, the sequence of links used, and the time of arrival. It is possible to imagine trip characteristics which contribute to the calculation of impedance, such as the number of stop signs and traffic signals encountered, the number of right turns, the number of left turns, and so on. In order to examine these aspects of automobile travel, it is necessary to examine the traffic assignment model in greater detail.

2.5.2 Traffic Assignment with TRANSIMS

In practice, the development of a traffic assignment model requires a considerable amount of ad-hoc adjustment. Moreover, these adjustments or calibration efforts can be performed along many dimensions. The link capacities, speeds and volume delay function parameters are all quantities specified by the analyst. In aggregate models, the number and configuration of centroid connectors constitute additional modifiable parameters. The demand matrix can also be altered to better fit observed flow patterns. These types of interventions are often performed manually and are poorly documented, if at all. The results of such efforts become difficult to justify and even more difficult to repeat, thus limiting the possibility of scientific experimentation. In the interests of clarity and transparency, a more systematic and rigorous method for constructing a useful model must be developed. The goal of this section is to demonstrate the potential of a totally disaggregate traffic assignment model to contribute to this evolution.

TRANSIMS is an open-source software package designed for performing activity-based transportation simulations, although its developers may have overlooked its greatest potential as a planning tool. It can be argued that the modelling of individuals at the microscopic level has little utility in the absence of real data which describe how people interact with their environment on a very small scale. Even if such data were available, the development of mathematical models to represent and reproduce the observed behaviour would require an almost unimaginable interdisciplinary effort. Among the many questions which would need to be addressed are: How is a vehicle shared among household members? How often do people go for groceries? How much do they buy? What size of vehicle do they need? How is the decision to go grocery shopping influenced by household structure? To what extent do people modify their travel behaviour to accommodate other household members? While it is unreasonable to expect that a model which accurately predicts these behavioural phenomena at the resolution of individuals and on the scale of an entire metropolitan could ever be constructed, it is possible that these questions are unimportant. A central concept in transport planning is the representation of "average" behaviour. The travel behaviour of individuals varies significantly from day to day, week to week and month to month. However, in the same way that the behaviour of a gas can be reliably predicted without accounting for the entirely random behaviour of its constituent molecules, it is not necessary to explicitly account for unpredictable variations in human behaviour when an estimate of a statistical mean is the only useful result. This is surely the case for forecast horizons which extend many years into the future.

From the perspective of the current study, TRANSIMS is especially noteworthy in two respects. First, it can generate large, detailed logistic networks, including control systems, automatically from easily obtained spatially-referenced databases of transport infrastructure. Second, it does not make use of an OD matrix but rather assigns trips from the point of origin to the point of destination. In this regard it is a totally disaggregate traffic assignment model. Because the assignment procedure conserves paths of individual travelers, the model is structured in such a way that it is possible to associate a household, its members and their attributes to particular road facilities through the simulated travel behaviour. The model is well-suited to application in the Greater Montreal Area since the region is periodically subject to a large sample origin-destination survey. The totally disaggregate structure of this travel demand database obviates the need for the population synthesizer and activity generator components of the TRANSIMS package.

The structure of the TRANSIMS traffic model is summarized in Figure 2.33. The model uses information describing transportation supply and demand and generates as output information on consumption. The representation of supply includes all the relevant components: nodes, links, movements and control systems. Vehicle trips ends are defined by the location of parking lots and person trips ends by the location of activities. Activities are connected to parking lots using artificial links which are somewhat analogous to centroid connectors in the traditional assignment model although they represent portions of the trip which are completed on foot. The representation of demand has the household-person-trip structure identical to the structure of travel information in the Montreal travel survey. Vehicles of various types are explicitly represented and are associated with particular trips, travellers, and households. The demand and supply information is used as input to a traffic simulator whose detailed functioning is described later. The simulator generates large quantities of output data which fully describe individual travel patterns (trip plans) as well as detailed indicators of network performance.

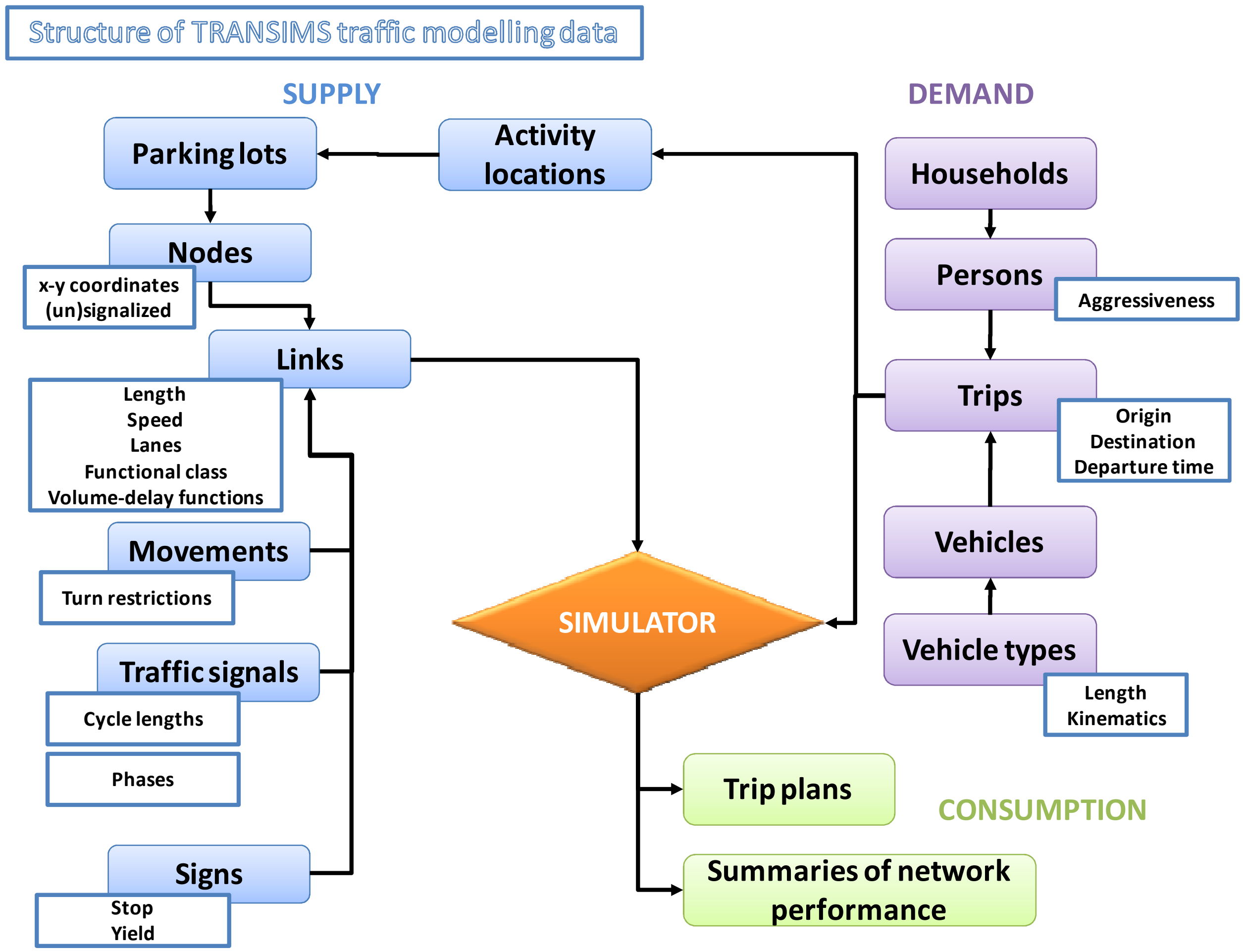


Figure 2.33: Structure of traffic simulation data in TRANSIMS

The goal of this section is to develop a simulation road network that reproduces, as closely as possible, the declared bridge choices of survey respondents. A network which reproduces the declared choices exactly would be a model of the road system as *perceived* by its users. A second requirement is that the network should be constructed using a minimal amount of manual intervention while maintaining a maximum amount of detail. The simulation network is a composite of two independent sources of data describing road infrastructure in detail. The first data source is fairly accurate in its representation of road geometry but provides a minimal amount of information on the attributes of its component links. The second network is coded to a high degree of geometric accuracy, its links are directed and it is designed for the purpose of transportation simulation. Both networks contain virtually every road segment in the Greater Montreal Area. The second network contains sufficient information to permit the synthesis of an artificial control system. The simulation network can therefore be adopted within a dynamic microsimulation framework or using a variant of the volume-delay method.

2.5.2.1 Representation of supply

TRANSIMS' path-finding module accepts input as data files in text or binary format. Interactive visual representation of input and output is accomplished through the use of the NEXTA visualization software (www.civil.utah.edu/~zhou/NEXTA_for_TRANSIMS.html) or TRANSIMS modules which convert the data files into .shp format readable by any GIS application. The two sources of network data used in the present model were summarily described in section 2.3.2. This section describes their integration into the TRANSIMS platform in more detail.

The GEOBASE representation is based on a digitized street network distributed freely by Natural Resources Canada through GEOBASE service (www.geobase.ca). This network contains almost all the streets in Greater Montreal, represented by 116,567 bidirectional links which connect at at-grade intersections. The network does not include a database of nodes. The geometry is accurate in most places (although there are some major inconsistencies) and each link is classified according to a functional hierarchy. The network which represents the Montreal region lacks important logistical information, particularly the designation of one-way streets and their orientation. Street names are not present either. The number of lanes is indicated, but since link directions are not specified, this attribute is of limited utility.

The Poly network is based on the road simulation network developed by Groupe MADITUC at École Polytechnique de Montréal., it includes only links. It represents almost every street in the Greater Montreal Area and each link has functional class and direction attributes. It is composed of 500 905 directional links. Like the GEOBASE network, there is no database of nodes. The large number of links is due to the fact that each link is a straight line with no vertices and the network is drawn in a way that is as faithful as possible to the real geometry. This network was subject to some considerable processing during its integration into the TRANSIMS platform. The pairs of directional links representing two-way streets were simplified to single bi-directional links, thereby reducing the number of links to 277 263. The TransimsNet module was then used to convert nodes which connect exactly two links into link vertices. The application of this module brings the total number of links to 104 227 and the total number of nodes to 69 896.

Although geometrically detailed, both networks lack information generally considered essential for microscopic simulation. To begin with, no information on the traffic control system (signals and signage) is included. Turning restrictions, other than those dictated by one-way streets, at intersections are absent. The number of lanes on each link and the lane configuration at intersections are excluded as well. Even if the necessary data were available, the incorporation of these elements would require more labour than is justified by a research experiment. Nevertheless, an effort was made to represent the 15 major bridges, in terms of number of lanes, free speed and traffic signal locations, as accurately as possible. Moreover, since the simulated demand constitutes only a small subsample of a 5% household survey, a complete representation of the network seems redundant.

It has already been shown (section 2.4.3.4) that the validated bridge-using trips in the travel survey comprise a scaled demand for travel on the major bridges. Also, section 2.4.2 demonstrated that bridge volumes obtained using the sum of trip record expansion factors (weights) were systematically lower than the volumes obtained by roadside counts and that the differences between the two measurements are attributable to the significant differences between the two data-collection methods. Rather than attempt to render compatible two independent sources of information by expanding sampled travel demand to the population, the supply is scaled to conform to the survey.

In a modelling framework governed by volume-delay functions, it is possible to scale the supply to match the demand since the travel time on a link depends entirely on the ratio of volume to capacity for a given time period. One method for scaling the bridge capacity is to assume that the bridge can accommodate the entire peak period demand over the entire duration of the peak period. In other words, without making any assumptions about the distribution of demand *during* the peak period, each bridge is assumed to be uncongested at 6 a.m. and is also uncongested at 9 a.m. Congestion occurs because of the non-uniform distribution of demand over the three hours. The three-hour capacity of facility j in direction a is therefore set equal to the number of declarations (N_j^a) in the travel survey corresponding to the use of facility j in direction a. Consequently, the hourly capacity of bridge j in direction a (c_j^a) is simply the number of responses during the entire three-hour period divided by 3. Inbound capacities (onto the Island of Montreal) are calculated using this method. Outbound capacities (exiting the Island) are computed based on the ratio of the number of inbound lanes (n_{in}) to the number of outbound lanes (n_{out}) since congestion in the outbound direction is assumed to be negligible. In other words:

$$c_j^{in} = \frac{N_j^{in}}{3} \qquad (2.15)$$

$$c_j^{out} = c_j^{in} \frac{n_j^{out}}{n_j^{in}} \qquad (2.16)$$

The static assignment also requires a free-flow speed for each link. The free-flow bridge speeds were based on the posted speed limits. The microsimulation requires that the number of lanes of a link be indicated explicitly. The complete list of the number of lanes, scaled capacities and specified free speeds is shown in Table 2.16 below.

Table 2.16: Specified bridge parameters for traffic assignment

BRIDGE	**LANES**		**SCALED CAPACITY (veh/h)**		**SPEED (km/h)**
	In	Out	In	Out	
CHAMPLAIN	3	2	207	138	70
VICTORIA	2	0	120	0	50
JACQUES-CARTIER	3	2	235	157	50
LAFONTAINE	2	2	165	165	70
MERCIER	2	2	151	151	70
GALIPEAULT	2	2	125	125	100
ILE-AUX-TOURTES	3	3	153	153	100
VIAU	2	2	69	69	50
PAPINEAU	3	3	149	149	100
PIE-IX	3	3	172	172	100
LACHAPELLE	3	3	61	61	50
MEDERIC-MARTIN	4	4	269	269	100
LOUIS-BISSON	4	3	280	210	100
LE GARDEUR	2	2	41	41	50
CHARLES-DE-GAULLE	3	3	183	183	100

In a microsimulation, network scaling is less straightforward. Link capacity is not an exogenous parameter but rather a result of a simulation incorporating dynamic queuing of individual vehicles. The length of a queue depends on the number of lanes and the lengths of its composite vehicles. Since the number of lanes is always a small integer, it cannot be scaled to the level of the survey (i.e. the number of lanes on a specific link cannot be divided by 25). In principle, the vehicle length is a more workable option for a scaling parameter. If, however, the sample of observed vehicles is small, as in the present case, the vehicle length can easily be on the order of hundreds of metres which, in addition to being unrealistic, is also longer than many of the links in a detailed network.

The primary interest of microsimulation in the present research is to explicitly simulate the control system, rather than attempt to derive a multitude of distinct volume delay functions to properly represent the urban street system. Usable data describing the signage and signals on Montreal-area streets were not available, so a control system was synthesized using the TransimsNet and IntControl modules in TRANSIMS. TransimsNet generates signal warrants for intersections according to a user-specified criterion. For this experiment, a traffic signal warrant was applied at all intersections where both roads had a functional class of “collector” or higher. The result was 4,446 signalized intersections out of a total of 69,896 nodes. Intersections without signals were assigned stop or yield signs depending on the intersection geometry. A sample of this synthesized control system is shown in Figure 2.34. Traffic signals are green dots, stop signs are red dots and yield signs are yellow dots. The IntControl module generates phasing plans based on a user-specified cycle length. A value of 90 seconds was used here, since it is the most common cycle length in the city of Montreal. The green time is apportioned between approaches based on the capacity of each.

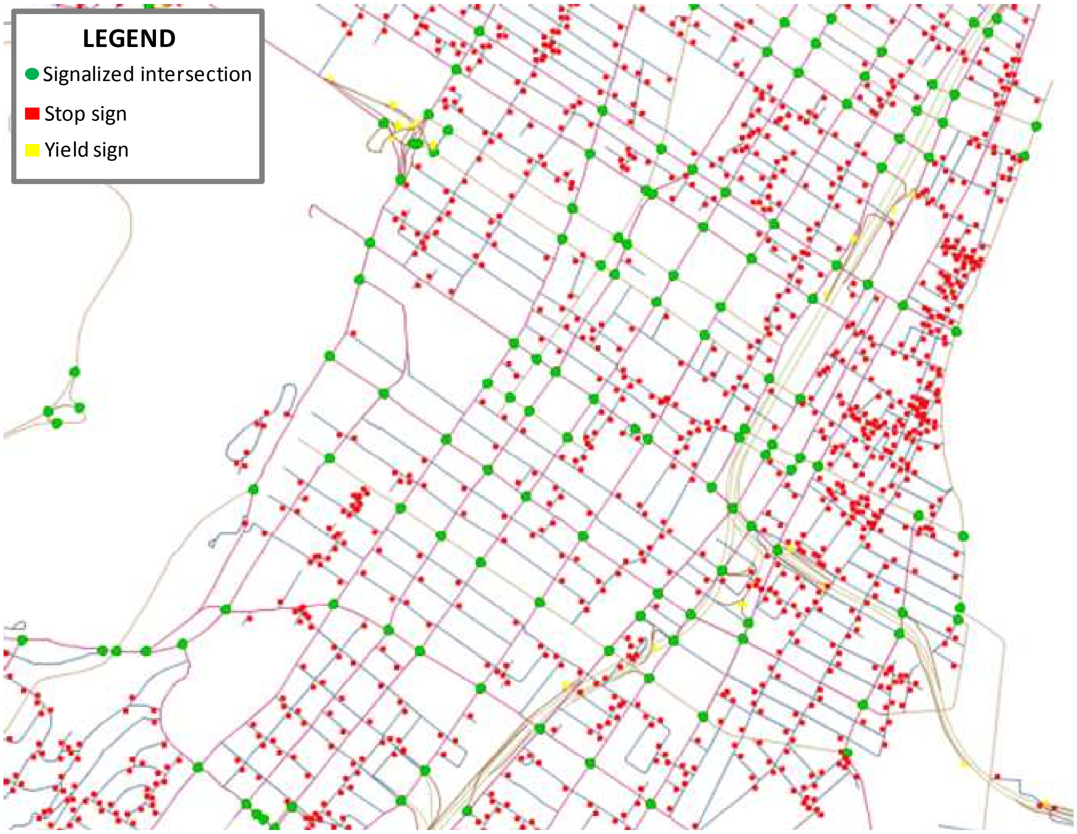


Figure 2.34: Downtown Montreal as represented by the simulation network.

2.5.2.2 Representation of demand

Demand is derived directly from the 8,583 trip records in the validated sample of the Montreal travel survey. Although TRANSIMS does not employ an origin-destination matrix and assigns trips between points in space rather than between zone centroids, it assumes that demand will initially be input as a matrix. The trip file is assumed to be produced through a synthesis process. The travel survey obviates the need for these steps and implies certain methodological adjustments.

TRANSIMS represents automobile demand using three separate files: trips, vehicles and vehicle type. The trip file structure is very similar to the structure of the Montreal travel survey database. Each trip has a point origin, a point destination and a departure time. Each traveller belongs to a specific household and is assigned a specific vehicle. Since TRANSIMS was conceived to model the activities of an individual constrained by interaction with other household members, the two most important trip attributes are the household number and the vehicle number. In the travel survey, each household has a unique identifier which is easily preserved in the TRANSIMS trip file. The survey also provides a unique identifier for each trip (called IPERE) linking the trip with its attributes, as well as the attributes of the traveller. It is therefore essential that the IPERE be preserved. TRANSIMS does not provide a unique trip identifier. The IPERE is therefore stored in the VEHICLE field. The same identifier is employed in the vehicle table. The vehicle type table allows for the definition of different vehicle classes, each defined by their length, their maximum speed and their maximum acceleration and deceleration rates.

2.5.2.3 Supply-Demand interactions

TRANSIMS simulates traffic through a feedback loop between an all-or-nothing assignment module (the Router) and a link performance calculator which computes average link travel times for time intervals of user-specified duration. The Router generates, for each trip, a trip plan (or itinerary) composed of a sequential list of links or nodes. The trip plan file is used as input to the link performance calculator. The complete simulation process is summarized in Figure 2.35.

Two options are available for computing volume-dependent link travel times. The first method employs a traditional volume-delay function using the LinkSum module. The second method is the simulation of individual vehicles using a cellular-automata model using the Microsimulator module. It is important to note that, regardless of which method is adopted, trips are routed through the network from the point of origin to the point of destination beginning at the indicated departure time. No origin-destination matrix is used. After the initial routing and link performance calculation, a third module (PlanCompare) compares the travel times of the initial plans with the travel times of the same plans based on the updated travel times. Trips with significantly different travel times are selected for re-routing in the next iteration. The itineraries of all other trips are conserved. At each iteration only a subset of trips is rerouted. The complete set of trips is considered in every link performance calculation. Ideally, the algorithm converges toward a stable solution. The stopping criterion is either a specified number of iterations or a number of rerouted trips which is deemed sufficiently small. An additional module (called LinkDelay) can be used to average link travel times between iterations, thereby attenuating the effect of shifted demand on travel cost and, in principle, improving the rate of convergence.

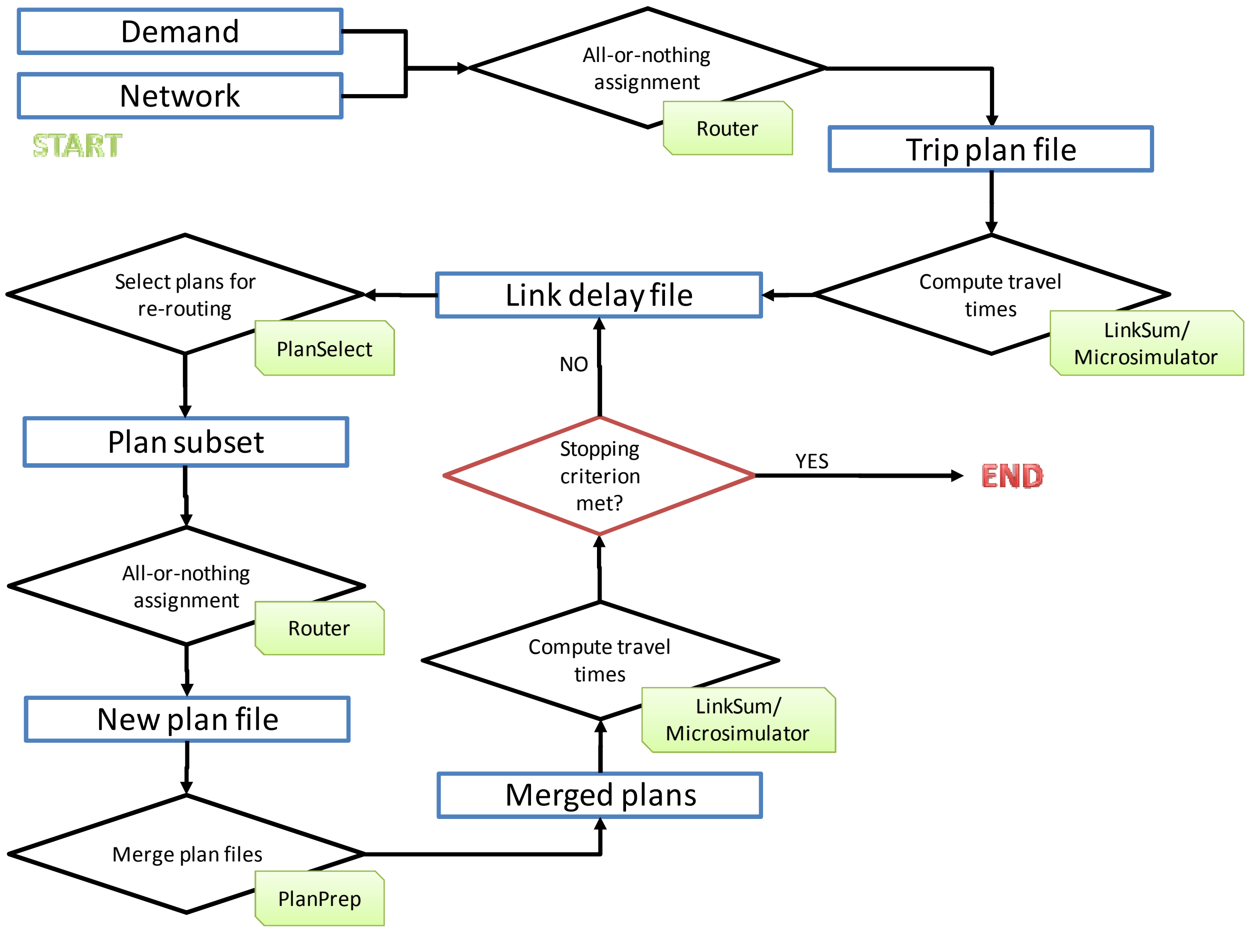


Figure 2.35: Summary of the TRANSIMS traffic assignment algorithm

The structure of the TRANSIMS traffic assignment means that microsimulation and volume-delay simulation methods are merely two alternative procedures for representing delay on a congested network. For the purposes of routing trips through the network, both the Microsimulator and LinkSum modules produce dynamic output: average travel costs per time interval. Apart from the algebra used to represent congestion, the primary difference between the macro- and micro-simulations is the input data. The volume-delay function, which can be applied globally or specified for a particular link, is composed of a free-flow travel-time, a link capacity and at least one calibration parameter. The microsimulator has no explicit volume-delay function but the relationship is represented by less artificial parameters. These variables include characteristics of the roadway (link length, number of lanes, free-flow speed, control system), characteristics of the vehicle (length, acceleration and deceleration rates, maximum speed) and characteristics of the driver (perception-reaction time, aggressiveness, lane-changing behaviour). It is not immediately obvious which of the two methods is preferable for the analysis of bridge choice. While the level of detail in a microsimulation is a more realistic representation of traffic – particularly with respect to the incorporation of traffic signals – it is also much more demanding in terms of the number of hypotheses which must be verified. The amount of coding is far greater than that which is required by the much more simplistic macro approach. According to the TRANSIMS documentation however, it is not necessary to choose one over the other. Both methods can be incorporated in a single simulation.

2.5.2.4 Model Output

The visualization of TRANSIMS results is independent of TRANSIMS itself. The software generates text and shapefiles as output. Both these formats can be easily read by a wide variety of common platforms. The NEXTA program generates graphics based on TRANSIMS text files, although it can be cumbersome when the network is large. TRANSIMS generates totally disaggregate simulation results which permit a detailed post-mortem of the traffic modelling process.

2.5.2.4.1 Paths

The complete list of links which make up a path are stored in a trip plan file. Each itinerary can be mapped. One pertinent application resulting from this arrangement is the rapid identification of facility users without resorting to the time-consuming “select-link” simulation. Figure 2.36 shows all the paths taken by users of the Pie-IX Bridge (circled in red). Note that, because trips are assigned point-to-point to a complete road network, the graphic is far more detailed than those typically generated by aggregate simulators. Given appropriate information, the duration and length of each itinerary can be compared to observed travel times to test the veracity of the model.

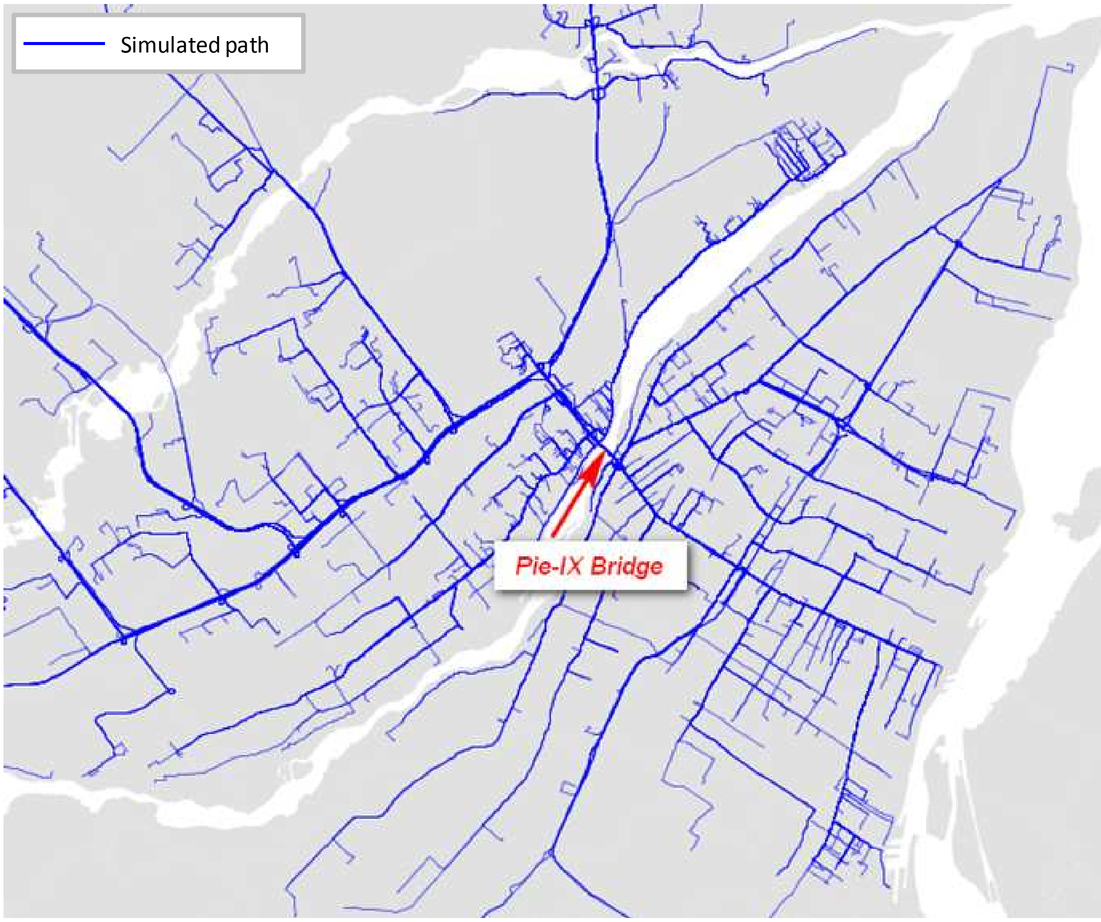


Figure 2.36: Paths taken by users of the Pie-IX Bridge

2.5.2.4.2 Link Performance

TRANSIMS computes link volumes, speeds, occupancies, queue lengths and measurements of delay, all of which are standard output in models of road traffic. Figure 2.37 shows an example of link flows generated by the assignment of the travel survey subsample to the detailed network. Figure 2.38 shows the link delays produced by a microsimulation. In the figure, the concept of delay is expressed as a ratio of the simulated congested travel time to the free-flow travel time. The travel time ratios are calculated for a fifteen minute interval (7:30 – 7:45) during the a.m. peak period. Traffic signals are represented by blue dots. In this case, delay is the ratio of the link travel time under free-flow conditions to the simulated link travel time. Since the microsimulation parameters were not scaled to match demand, the major bridges and freeways experience no delay. The delay visible on the urban road network is entirely due to the presence of traffic signals.

The differentiation of results by time interval allows for the construction time-varying link speeds (Figure 2.43) and, by extension, a network “schedule”. In theory, this schedule could be provided as an exogenous input based on direct observation. In practice, the schedule is constructed by computing link costs using either a volume-delay function or a microsimulator. In the models presented below, 15-minute time intervals are used to calculate average link performance statistics.

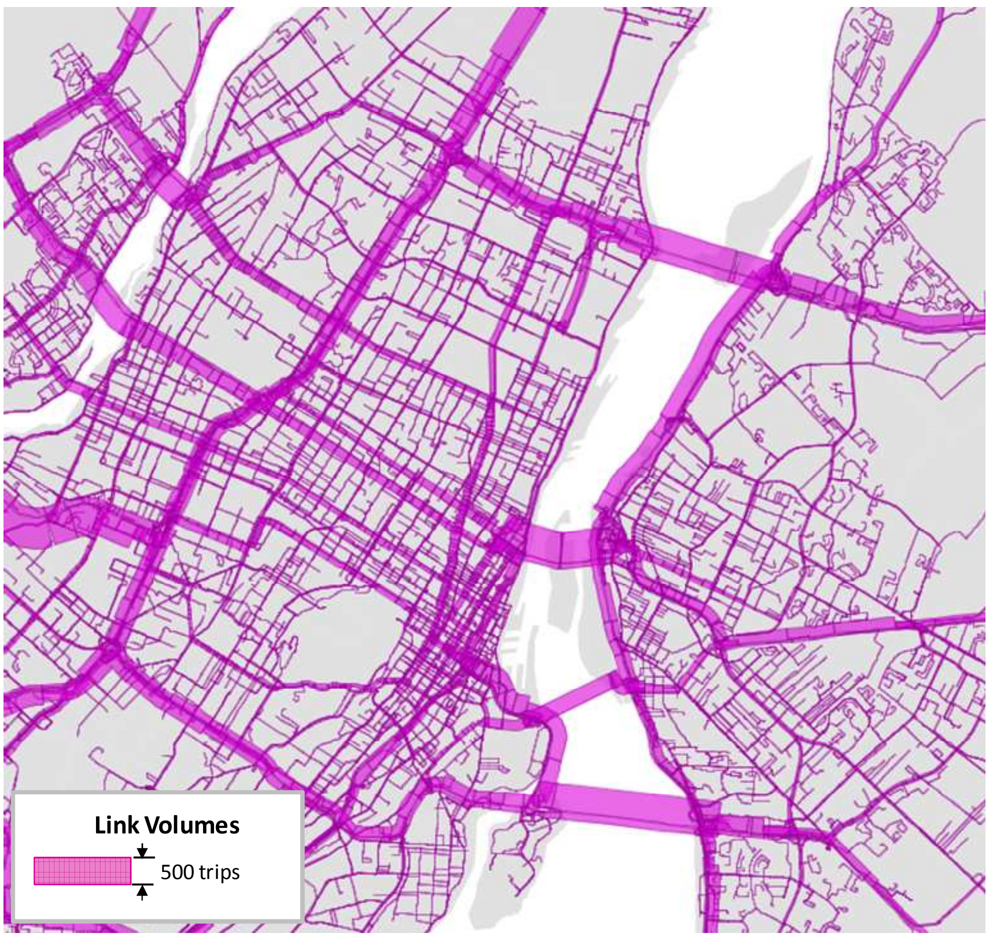


Figure 2.37: Flow map produced using a totally disaggregate assignment of bridge traffic for the a.m. peak period

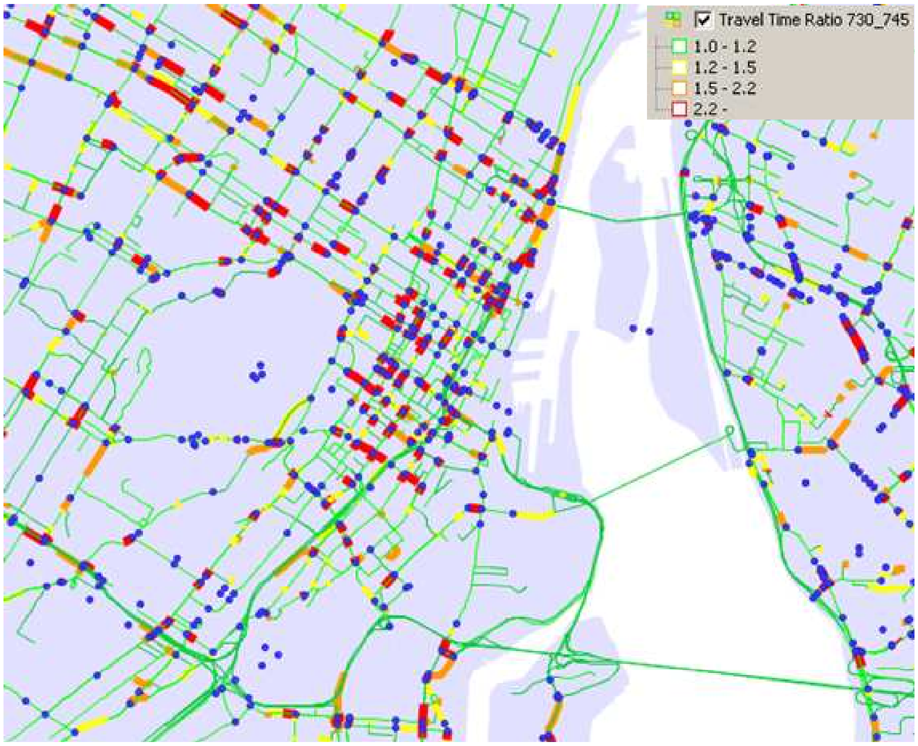


Figure 2.38: Link delay 7:30-7:45 as generated by a microsimulation.

2.5.2.5 Analysis of itineraries

The output of the TRANSIMS simulation allows for the close examination of individual trips. Such an exercise provides some insight into the micro-level factors that influence the choice of route and the choice of bridge. These factors can be detected by comparing the predicted itinerary of a trip with the itinerary corresponding to the information provided by the traveller in cases where these two itineraries are not the same. In other words, incorrect predictions of the model are used to gain insight into traveller behaviour.

For the purposes of illustration, the detailed analysis of an itinerary is based on the simulation and validation models described in section 2.4.4.1. These models are simple all-or-nothing assignments without congestion or traffic signals. One trip was chosen from the 133 trips that were assigned to the Jacques-Cartier Bridge but in reality used the Champlain (see Table 2.8). The trip originates in the suburban community of St-Hubert (Longueuil) and is destined to a location east of downtown Montreal. The traveller is a 28 year-old female leaving her home at 6:00 in the morning to go to work. Figure 2.39 shows, for the selected trip, the itineraries generated by the simulation and validation models. A detailed comparison of the two itineraries suggests reasons why the traveller chose a route different from the one proposed by the model.

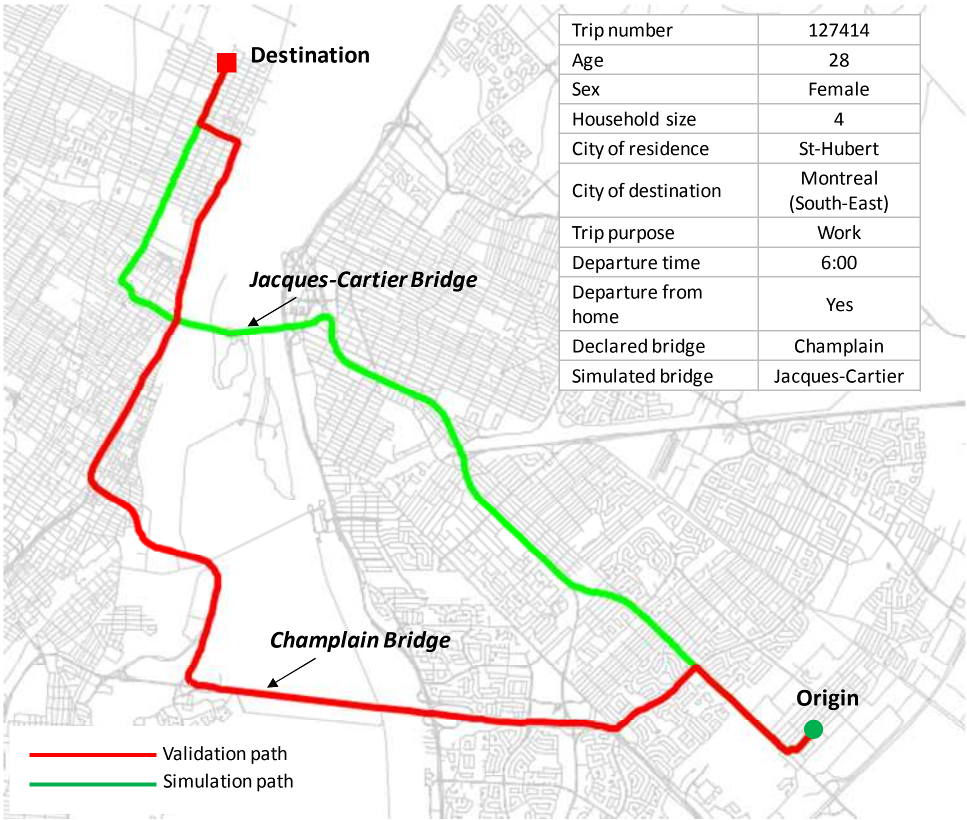


Figure 2.39: Comparison of simulated and validated itineraries for a single trip

A summary comparison of the two itineraries is shown in Table 2.17. The table shows the distributions of distance and time over the different functional road classes, as well as a compilation of all the turning movements required to complete the trip. The validation itinerary corresponding to the declared bridge choice is 5.5 kilometres (23%) longer than the simulated itinerary. The travel times for the two routes are comparable; 22.0 minutes for the validation itinerary and 19.8 for the simulation itinerary. This means that the route chosen by the trip-maker allows for travel at a higher speed than the route chosen by the model. Indeed, the validation itinerary includes 12.2 km of freeway travel as opposed to 3.5 km of freeway in the simulation path and the respective average speeds of the two itineraries are 63 km/h and 56 km/h.

The validation path requires 12 turning movements while the simulation path requires only 6. Seven of the manoeuvres in the validation path are right-hand splits and merges which, from the driver's perspective, are theoretically less demanding in terms of time, attention and effort than right and left turns. In addition, the simulation path requires two merges from the left – a movement which is unusual on North American road networks. The quantities of thru movements are worth comparing also. The validation path includes 71 at-grade intersections, including freeway entrances and exits. By contrast, the simulation path involves 105 at-grade intersections. Although true location of traffic signals along each route is unknown, it is very likely that a driver following the simulation itinerary using the Jacques-Cartier Bridge will encounter considerably more red lights than if she chooses the Champlain Bridge.

Table 2.17: Summary statistics for the simulated and validated itineraries

	VALIDATED	**SIMULATED**
Distance (km)		
Local	0.8	0.7
Arterial	9.5	13.8
Ramp	0.6	0.6
Freeway	12.2	3.5
TOTAL	**23.1**	**18.6**
Time (min.)		
Local	1.2	1.0
Arterial	9.9	15.1
Ramp	0.8	0.7
Freeway	10.1	3.0
TOTAL	**22.0**	**19.8**
Average speed (km/h)	**63.0**	**56.2**
Movements		
Thru	71	105
Right split	4	0
Left split	0	0
Right merge	3	0
Left merge	0	2
Right turn	2	3
Left turn	3	1
TOTAL	**83**	**111**

Table 2.18 lists the turning movements of both itineraries in chronological order. The time and distance elapsed between movements is included. Initially, the validation itinerary seems much more complicated than the simulation path. Twice as many movements are required and in several cases the driver has 15 seconds or less between manoeuvres. Most of the movements in the validation path, however, are simple merges and splits when entering and exiting the freeway. Left and right turns are required only at the very beginning and the very end of the trip. The simulation itinerary, meanwhile, requires right turn just 11 seconds after a merge from the left. The importance of these microscopic phenomena in the driver's choice of bridge merits

further study. More network details such as lane configurations and turning restrictions are required for a more extensive investigation.

Table 2.18: Detailed descriptions of the simulated and validated itineraries

VALIDATED ITINERARY				SIMULATED ITINERARY			
Movement	Time	Elapsed time	Elapsed distance (km)	Movement	Time	Elapsed time	Elapsed distance (km)
LEFT	6:00:51	0:00:51	0.6	LEFT	6:00:51	0:00:51	0.6
RIGHT	6:01:01	0:00:10	0.1	RIGHT	6:01:01	0:00:10	0.1
LEFT	6:02:59	0:01:58	2.0	RIGHT	6:05:26	0:04:25	4.4
R_SPLIT	6:04:26	0:01:26	1.4	L_MERGE	6:11:04	0:05:38	6.1
R_SPLIT	6:04:38	0:00:12	0.1	L_MERGE	6:15:00	0:03:56	3.4
R_MERGE	6:04:53	0:00:15	0.2	RIGHT	6:15:11	0:00:11	0.2
R_MERGE	6:06:28	0:01:35	1.3				
R_SPLIT	6:10:48	0:04:20	5.2				
R_SPLIT	6:14:22	0:03:34	4.1				
R_MERGE	6:14:54	0:00:32	0.7				
LEFT	6:20:03	0:05:09	5.7				
RIGHT	6:20:43	0:00:40	0.7				

Much of the above information can be summarized in a space-time diagram (Figure 2.40). The dotted lines represent the vehicle trajectories in space. The hollow circles or squares indicate positions where a turning movement is required. The series of connected squares indicate the vehicle speed as a function of time. The most obvious difference between the two routes is the higher speeds experienced on the validation path on the Montreal side of the bridge. A vehicle following the Champlain Bridge can continue to travel at 70 km/h or 60 km/h until the final minute of the journey. If the driver chooses the Jacques-Cartier, he is forced to travel 50 km/h for the entire portion of the trip from the bridge exit to the destination. These results are based on a simulation without congestion. The effects of congestion on the choice of itinerary during the access and egress portions of the trips are unclear in the absence of additional observed route information. As usual, the simulated traffic speeds are likely much higher than speeds on the network but any adjustments in the absence of direct speed measurements would be arbitrary.

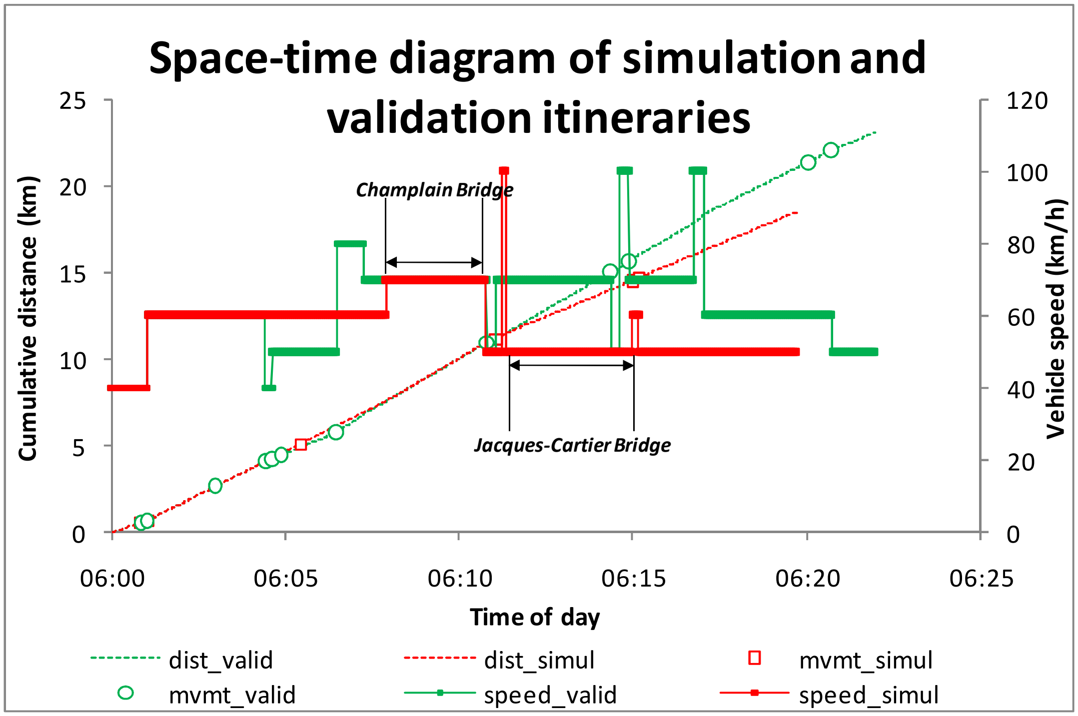


Figure 2.40: Space-time diagram of the validation and simulation itineraries

2.5.2.6 Incorporation of congestion effects

It has already been shown that an all-or-nothing assignment is capable of reproducing about three quarters of observed bridge choices. The explanatory power of this simple model was not significantly improved by the incorporation of traffic congestion using unconventional methods. In section 2.5.1.1, a discrete choice model was estimated that accounted for congestion through utilization ratio applied to the major bridges. Another model used congested travel times on network links to compute the time cost of each alternative route. Here, a more standard approach is adopted. Travel time on every network link is assumed to be dependent on traffic volume. The relationship between volume and travel time can be expressed using a volume-delay function or through the explicit representation of queues (microsimulation). In either case, the disaggregate structure of the demand data must be maintained.

TRANSIMS provides a tool for performing iterative capacity-constrained assignment of individual trips. A ten-iteration simulation algorithm is summarized in Figure 2.41 below. The process is initiated with an all-or-nothing assignment of travel demand to the minimum-time path on the network. A delay mechanism is used to recompute link travel times. During the first five iterations, link travel times are computed using a volume-delay function applied to all links that have a non-local functional class. Local streets are assumed to be uncongested. Although the model is not microscopic, it is not static either. Even when a volume-delay function is used to represent congestion, TRANSIMS recalculates link travel times at regular simulation time intervals meaning that travel times change over the course of the simulation period just like a real road network. Fifteen-minute intervals are used in the simulations described below. A standard BPR function is adopted for the calculation of t_m, the link travel time during time interval m:

$$t_m = t_0\left[1 + 0.15\left(\frac{V_m}{C}\right)^4\right] \qquad (2.17)$$

Where t_0 is the link travel time under free-flow conditions, V_m is the link volume during time interval m, and C is the link capacity.

The last five iterations consist of a microsimulation which represents traffic signals and queuing effects explicitly. This hybrid approach is recommended in the TRANSIMS documentation (see, for example, http://code.google.com/p/transims/wiki/GettingStartedRun). At a given iteration, the link travel times from the previous iteration are used to determine the shortest path for each trip. If, for a given journey, the travel time calculated in the previous iteration exceeds the new travel time by an amount less than ϵ, then the trip is not rerouted. Its assigned route from the previous iteration is used to compute link flows.

When travel times are calculated using a volume-delay function, it is possible to scale the supply to match the 5% sample of demand (see section 2.5.2.1 for details). In this way, congestion effects are generated on the major bridges. Because the road supply cannot be scaled during microsimulation, congestion develops primarily at traffic signals on the urban network. Congestion effects are therefore less important during the microsimulation. In consequence, a different value of ϵ is used for the two delay models. In the volume-delay model, only trips having an ϵ at least 50% greater than the travel time from the previous iteration are re-routed. During microsimulation, trips having ϵ at least 10% greater than the previous travel time are re-routed.

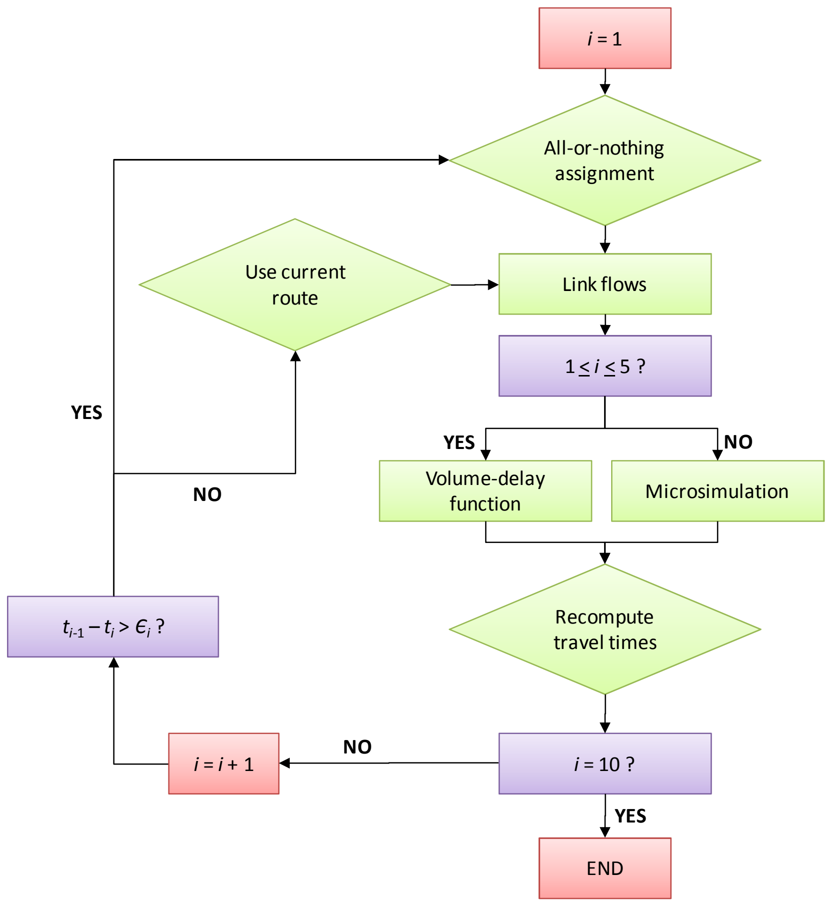


Figure 2.41: TRANSIMS assignment/route choice procedure

The goal of this research is to estimate consumption levels based on the observed usage patterns of major road facilities (bridges) which are the aggregate result of many individual choices. It is therefore important to examine the effects of iterative simulation and congestion effects on the model's ability to reproduce observed choices. Figure 2.42 shows the evolution of the correct prediction rate over the 10 iterations of the simulation for selected bridges and for all the bridges together. Also shown is the number of trips re-routed between iterations. Some of the new routes result in a change of bridge. The number of trips that change bridge between iterations is indicated as well.

Note that the total correct prediction rate (the black dotted line) hardly changes throughout the simulation. The correct prediction rates of individual bridges do fluctuate but an improvement in the performance of one bridge usually comes at the direct expense of another (1303 vs. 1302 and 1401 vs. 1402). Also, although the number of rerouted trips frequently exceeds 3000, the number of trips which change bridge between iterations never reaches 1000. This means that, for many travellers, the increase in travel time due to congestion is still less than the excess travel time that would result from the use of an alternative bridge. In summary, the partial incorporation of congestion into the model, despite having a significant influence on the choice of bridge, does not increase the total number of correct choice predictions.

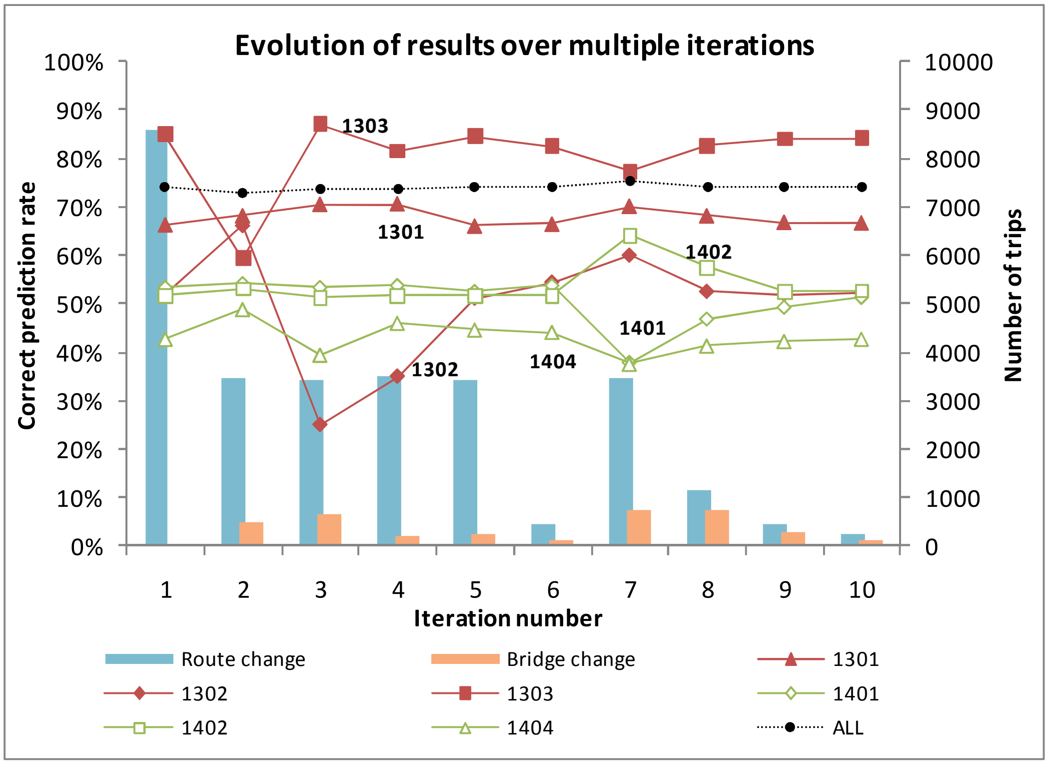


Figure 2.42: Evolution of assignment results over multiple iterations

Although volume-delay function employed in the first 5 iterations was not calibrated, the scaled bridge capacities produce significant congestion effects. An example of these is shown in Figure 2.43 which represents the relationship between volume and average link speed on the five bridges of the South Shore screenline in the inbound direction over the course of the simulation period. The results are taken from iteration 5. Four of the bridges experience significant reductions in speed. The average speed of the Victoria Bridge remains stable since it attracts a level of demand which is well-below its capacity. The relative unattractiveness of the Victoria Bridge is an unrealistic feature shared by the all the facility choice models, including those based on random utility. The difficulty in obtaining a representative result for the Victoria Bridge may be attributed to its position roughly midway between the Champlain and Jacques-Cartier Bridges which, like the Victoria, provide access from the South Shore to the downtown core. The coding of the street network at either end of the bridge may also play a role.

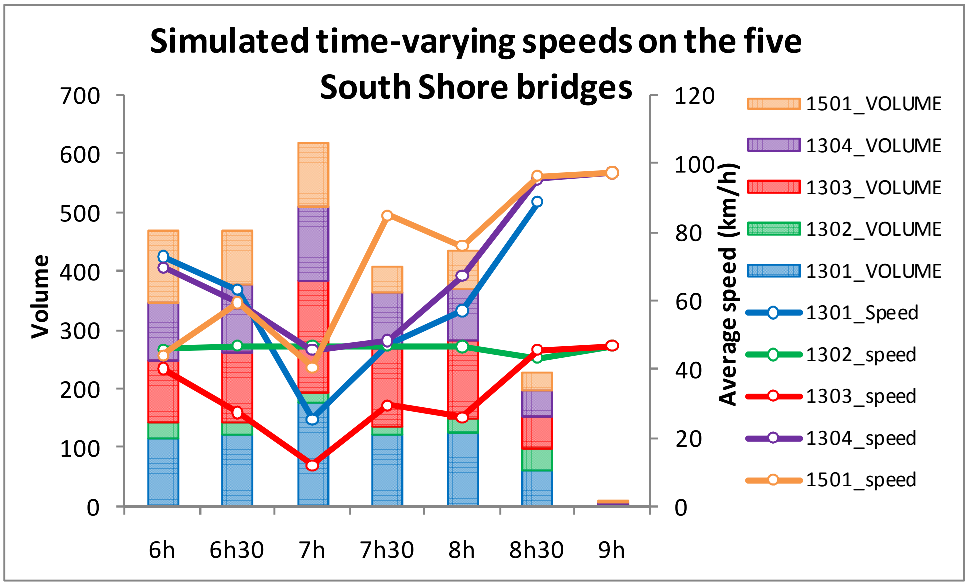


Figure 2.43: Time-varying link speeds and volumes on the five major bridges of the inbound South Shore screenline simulated using volume-delay functions.

The amount of simulated congestion can also be assessed through an examination of the global evolution of travel costs over the course of the peak period, as shown in Figure 2.44. Average trip length is computed as the total vehicle-kilometres travelled divided by the number of vehicles and average speed is the total vehicle-kilometres divided by the total vehicle-hours. The figure shows the simulated variation in average trip lengths and speeds over the course of the a.m. peak period. The results of iteration 1 and iteration 10 are both shown. Although the difference is small, average speeds are lower for each half-hour period in the 10th iteration than in the first iteration, indicating the imposition of some congestion effects. Average trip lengths (in kilometres) also become longer between the first and last iteration. In addition, the average trip length decreases with respect to time of day which suggests the existence of a universally preferred arrival time; people making long trips must leave earlier in order to arrive around the same time as the people making shorter trips. This phenomenon is discussed further in Chapter 3.

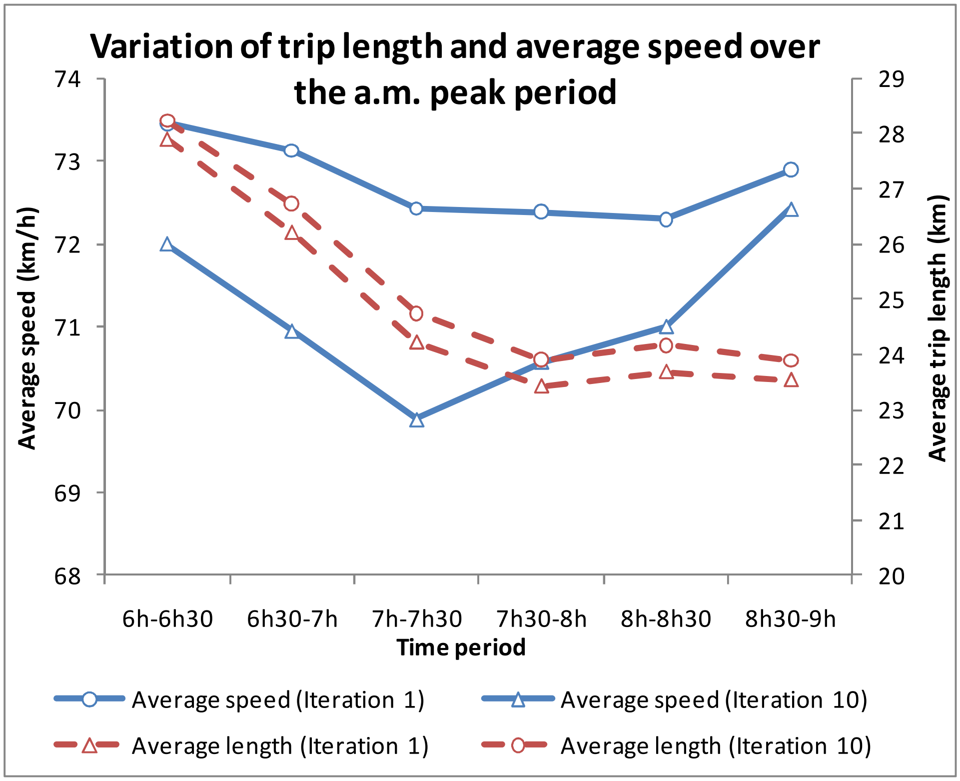


Figure 2.44: Simulated evolution of average network speed and average trip length over the a.m. peak period

This simulation is meant to reproduce congestion on bridges and freeways using a conventional volume-delay function, and then re-route some trips to account for the presence of traffic signals. Despite the considerable additional modelling effort – the incorporation of signals, the execution of an iterative algorithm – the result with respect to bridge choice is no better than the one obtained using an all-or-nothing assignment. The confusion matrix for bridge predictions at the 10th iteration (Table 2.19) shows a correct prediction rate of 74.1% that is identical to the rate obtained at the end of the first iteration (the all-or-nothing assignment the results of which are shown in Table 2.8).

Table 2.19: Confusion matrix for the iterative model (result at the end of the 10th iteration)

Number of trips		Modeled bridge																
Screenline	**Observed bridge**	1301	1302	1303	1304	1501	1401	1402	1403	1404	1405	1406	1503	1504	1601	1602	TOTAL	% CORRECT
South Shore	1301	519	84	130	21	24											778	66.7%
	1302	46	188	116	7	3											360	52.2%
	1303	36	71	696	23	2											828	84.1%
	1304	2	2	127	583												714	81.7%
	1501	23		4	4	512							4	2			549	93.3%
Laval	1401						125	55	13	2	41	5			1	1	243	51.4%
	1402						31	295	132		96	2			1	3	560	52.7%
	1403						9	82	504		47	3				9	654	77.1%
	1404						14	10	2	91	77	18				1	213	42.7%
	1405						45	99	66	29	714	64				1	1018	70.1%
	1406						6	27	21	67	173	711			1	1	1007	70.6%
West	1503												384	24			408	94.1%
	1504					1							46	464			511	90.8%
East	1601														37	98	135	27.4%
	1602							1	55		2	1			9	537	605	88.8%
	TOTAL	626	345	1073	638	542	230	569	793	189	1150	804	434	490	49	651	8583	**74.1%**
	% ERROR	-20%	-4%	30%	-11%	-1%	-5%	2%	21%	-11%	13%	-20%	6%	-4%	-64%	8%	**86.7%**	**19.7%**
																	R^2	RMSE

2.6 Conclusions

2.6.1 Validation of revealed preference information

Section 2.4 examined in detail revealed preference information concerning the choice of major bridge contained in a large-sample telephone travel survey. The results of the exploratory data analysis and the subsequent estimation of models based on a validated subsample suggest that the survey responses are, on the whole, coherent and sensible. Gross errors attributable to the data collection and codification processes were believed to comprise less than 5% of the validated sample. The information is sufficiently credible to be used as the basis for a detailed analysis of infrastructure usage patterns.

2.6.2 Demonstration of the methodology

This chapter has described the construction and application of two types of facility choice model: a random-utility model and a traffic assignment method. Both model structures adopt microscopic, dynamic and totally disaggregate representations of supply and demand. All the demonstrated models were evaluated on their ability to reproduce choices observed in a revealed-preference survey of travel behaviour. All the models were able to correctly predict around 75% of the observed behaviour. The procedure demonstrates the feasibility of developing detailed representations of supply and demand with a minimum of human intervention, thus saving time and facilitating the application of a scientific approach. It also illustrates how the behaviour of automobile drivers on an urban network can be studied using totally disaggregate methods which incorporate accepted theories of travel behaviour while remaining independent of the classic four-stage approach. The disaggregate data structure means not only that certain travel behaviour can be represented in more detail, but also that the workings of the model itself can be thoroughly examined using, for example, a confusion matrix.

2.6.3 Comparisons between models

A detailed summary of the modelling results, in terms of the correct prediction rate, is shown in Table 2.20. The question of which model best represents each bridge is of secondary importance since no significant effort was expended on the verification and correction of any network. For just one regional model, the task is monumental and often interminable. Such an exercise could be undertaken for each of the 5 models tested. Their absolute and relative predictive powers would presumably change as a result. What is notable about Table 2.20, however, is the *stability* of the modeling outcomes. The total correct prediction for each model hardly varies. The worst performing model is the iterative assignment and it successfully predicts 74.1%. The model with the most predictive power is the logit model using congested travel times and its correct prediction rate is 75.7%. The variation in model accuracy is greater at the level of individual bridges. Bridges that do not carry freeways (1302, 1401, 1404 and 1602) are consistently poorly represented. The fact that the incorporation of congestion effects tends to improve the usage forecasts of these facilities suggests the existence of the classic user-equilibrium under congested conditions: facilities that are usually slow become reasonable options when the freeway facilities are congested.

Another conclusion that can be drawn from this modelling exercise is that in many cases the prediction of bridge choice is almost trivial. If this conclusion could be demonstrated to extend to the choice of all major road facilities, then it would constitute an encouraging result: most of the users of a particular facility can be determined using simple methods. Simultaneous correct prediction of the usage of multiple facilities is certainly more difficult. If, for example, 75% of the users of a single bridge can be properly represented, then it is likely that the same modelling framework would reproduce the real behaviour of only 56% (0.75 * 0.75) of travellers who use two bridges. In many planning applications, however, the analysis of the actual and potential clientele of a single facility is often all that is necessary.

Table 2.20: Summary of model performance using the correct prediction rate

Method	**Assignment**		**Discrete Choice**		
	Freeflow	Iterative	Logit 1	Logit 2	Logit 3
Bridge					
1301	66.3%	66.7%	85.7%	85.9%	78.4%
1302	51.7%	52.2%	39.4%	45.3%	43.1%
1303	84.9%	84.1%	69.9%	62.8%	76.9%
1304	81.4%	81.7%	86.7%	88.8%	86.7%
1501	93.3%	93.3%	90.2%	89.4%	94.7%
1401	53.5%	51.4%	48.6%	52.3%	59.3%
1402	51.6%	52.7%	50.9%	56.8%	61.4%
1403	78.6%	77.1%	78.3%	74.0%	67.1%
1404	42.7%	42.7%	26.8%	33.8%	37.1%
1405	69.7%	70.1%	76.5%	74.2%	65.2%
1406	70.4%	70.6%	75.0%	78.9%	84.1%
1503	94.1%	94.1%	94.6%	94.6%	90.4%
1504	90.8%	90.8%	91.2%	91.6%	94.7%
1601	27.4%	27.4%	12.6%	14.8%	17.0%
1602	88.8%	88.8%	91.6%	92.6%	93.2%
TOTAL	**74.1%**	**74.1%**	**74.9%**	**75.3%**	**75.7%**

2.6.4 Model Complexity and Explanatory Power

Another aspect of this investigation concerns the prospects of improving the explanatory power of a behavioural model. Very often, the incorrect predictions of models are attributed to missing explanatory variables. The literature is full of instances where models are extended to account for previously ignored phenomena. While this approach has merit in particular situations, the present research has illustrated a case where additional complexity adds little or nothing to the quality of the result. Table 2.20 is offered in support of this assertion. When the correct prediction rate of a given bridge increases from one model to another, there is necessarily another bridge whose correct prediction rate decreases. Stated differently, there are always some travellers whose behaviour is correctly represented by one model but is poorly represented in another, even if it is more sophisticated. From this perspective, the challenge is not to increase the complexity of the model, but rather to better understand the individual-specific context in which each person makes decisions. More generally, this result should not be surprising when the behaviour being modelled concerns the choice of route. It is well-known that the optimization of traffic flow through a network constitutes a non-convex problem with respect to paths, meaning that there exists a multitude of equivalent routing solutions. Since only one of these solutions will correspond to one observed reality, it is natural that any route choice model will have limited predictive power.

2.6.5 The role of congestion in the modelling of bridge choice

In the experiments described above, congestion was represented using several alternative methods. In section 2.5.1, congestion was represented in logit Model 2 by a utilization ratio computed as the volume predicted by logit Model 1 divided by the observed volume. In logit Model 3, congestion was incorporated as an exogenous input based on direct observation. In section 2.5.2, congestion effects were incorporated using both a volume-delay function and a microsimulation to generate queues at traffic signals. None of these models are methodologically correct representations of the congestion on major bridges. The queues which develop upstream of each major bridge would ideally be modelled through microsimulation using either scaled supply (for which no workable method could be found) or the complete demand (the survey is only a 5% sample). The fact remains, however, that the predictive power of the models which attempt to account for traffic congestion is not significantly greater than the predictive power of models which do not.

In this thesis, the term traffic congestion refers to recurrent congestion only. At issue, then, is the difference between travel times under free-flow conditions and *average* travel times during periods of peak demand. It is not difficult to believe that a change in average travel times exerts little influence over the choice of congestible facility. Congestion occurs, after all, because there is a lack of alternatives (choice). It seems likely, however, that facility choice is influenced by non-recurrent incidents which travellers learn about through traffic reports. In the case of Montreal, traffic updates on the radio often include estimated waiting times to access the major bridges. This awareness of the short-term variations in network conditions could lead to a queue-jockeying effect (Koenigsberg, 1966) where travellers alter their choice of bridge during the course of the trip or tend to choose a different bridge each day. Alternative modelling methods exist to represent this phenomenon. In this paradigm, traffic on a network can be considered a "prisoner's dilemma" where the availability of only partial information leads decision-makers to make a sub-optimal choice. There is an evident connection with Braess' paradox (Hassin & Haviv, 2003). Finally, the rapid evolution of wireless technology and in-vehicle navigation systems offers intriguing possibilities for providing more detailed information to drivers. It remains to be seen exactly how these user-information systems will influence driver route and facility choice.

Although the results of this modelling process suggest that the consideration of traffic congestion has a limited impact on the ability of a model to predict the choice of bridge, there is no doubt that congestion influences travel times. While many methods exist to simulate congestion, additional data are required to determine how they can best be applied in a predictive model. Road networks follow a schedule and this schedule must be observed and analysed if a coherent model of travel time by automobile is to be constructed. This objective will be the topic of future research.

2.6.6 Implications for planning

This chapter has described a method for approaching the route-choice/traffic assignment problem which incorporates partial information on the chosen paths. The experiment has demonstrated that partial representations of supply and demand and simple hypotheses about driver behaviour are sufficient to reproduce around 75% of the usage patterns of 15 congested road facilities. This finding seems especially pertinent to the analysis of proposed infrastructure projects, which typically impact only a limited portion of the network and a particular segment of the travel demand market. It has also been shown that, even though the facilities may be subject to congestion, the phenomenon does not need to be explicitly represented in order to achieve a high correct prediction rate. Note that comparisons of predicted flows with observed volumes are not at issue here. A determination coefficient was computed for each model and it exceeds 90% - a very respectable score for a standard traffic assignment – in almost all cases. It has not been discussed as a measure of model performance since it says nothing about how well the model reproduces *individual choices*. The determination coefficient is an aggregate result which masks incorrect predictions.

CHAPTER 3 ANALYSIS OF EQUITABLE ROAD TRANSPORT IN THE GREATER MONTREAL AREA

The equitable distribution of transport costs and benefits within an urban area is a problem of considerable importance given the current responsibilities of highway planners and engineers. Urban sprawl, greenhouse gas emissions, oil consumption and the imperative of sustainable development are issues which require a thorough understanding of the interaction of the road transport system with the people it serves. The nature of this interaction is the result of transportation policies designed to meet particular objectives, many of which have little to do with transportation itself. These policies nevertheless stimulate certain types of behaviour while discouraging others. If conscious intervention in the operation of urban road systems for the purposes of achieving societal goals is to be a meaningful exercise, then the effects of proposed policies must be quantifiable and, to some extent, predictable. This chapter outlines a method by which the goal of equity (fairness) can be quantified using observed travel patterns and demonstrates that comparable results can be obtained through the application of an appropriate simulation model.

This research is by no means the first effort to discuss equitable road transport and to quantify the redistributive effects of a network using the totally disaggregate modelling approach. Issues surrounding the equitable distribution of travel costs and benefits have been of interest to researchers at École Polytechnique de Montréal for many years (see (Chapleau, 1988; Bergeron & Chapleau, 1996; Essakali, 1999; Chapleau & Morency, 2004) among others). While the following demonstration is largely based on the approach adopted in these earlier projects, it is focused on major components of road infrastructure, particularly the fifteen bridges of Montreal. A study of such narrow scope is justified by the fact that specific public works, such as bridges, freeways, metros and so forth, exert a strong influence over travel patterns throughout an urban area. Indeed, the bridges of Montreal play a fundamental role in the distribution of transport-related costs and benefits over space and time within the Greater Montreal Area. This chapter will demonstrate how the redistributive effects generated by particular infrastructure elements can be credibly measured and how the analysis results can be used to construct an equitable road pricing system. The methodology is theoretically extendable to any major road infrastructure components whose usage patterns can be observed.

Figure 3.1 is a conceptual diagram of an equity analysis for major road infrastructure in an urban environment and summarizes the methodology described in this chapter. The existence of the major bridges induces a certain amount of personal travel by car. The automobile trips accomplished by these drivers constitute movement over the road network. This interaction of transport demand (trips) with transport supply (the network) is represented by a traffic assignment model of the type described in detail in the previous chapter. Initially, the analysis is performed using the validation model in order to quantify various phenomena using collected information. The analysis is then repeated using a simulation model to test the sensitivity of various indicators to the simplifying hypotheses of traffic model. In both cases, the result of the assignment process is a link-by-link itinerary for each trip. Each link in an itinerary can be classified according to the jurisdictional or functional network to which it belongs. The functional hierarchy is used to estimate the costs and benefits of travel which accrue to the individual driver. The jurisdictional hierarchy determines how costs are distributed among geopolitical entities. The costs which are borne by territorial governments are passed on to the households located within those territories through taxation. Households bear additional non-monetary costs based on their location in space. At the level of geopolitical entities, a comparison of the costs and benefits incurred by each jurisdiction permits the calculation of equity indicators.

This chapter has eight sections. The first section of this chapter discusses the philosophical notion of equity within the context of transportation engineering. The second section describes the costs and benefits of road transport and how they might be measured. The third section deals with the redistributive effects of major road infrastructure within an urban region. Section 3.4 outlines a quantitative methodology, section 3.5 reveals the results of the analysis using the validation model, section 3.6 repeats the analysis using the simulation model, section 3.7 re-examines traffic models in the context of these findings and section 3.8 concludes.

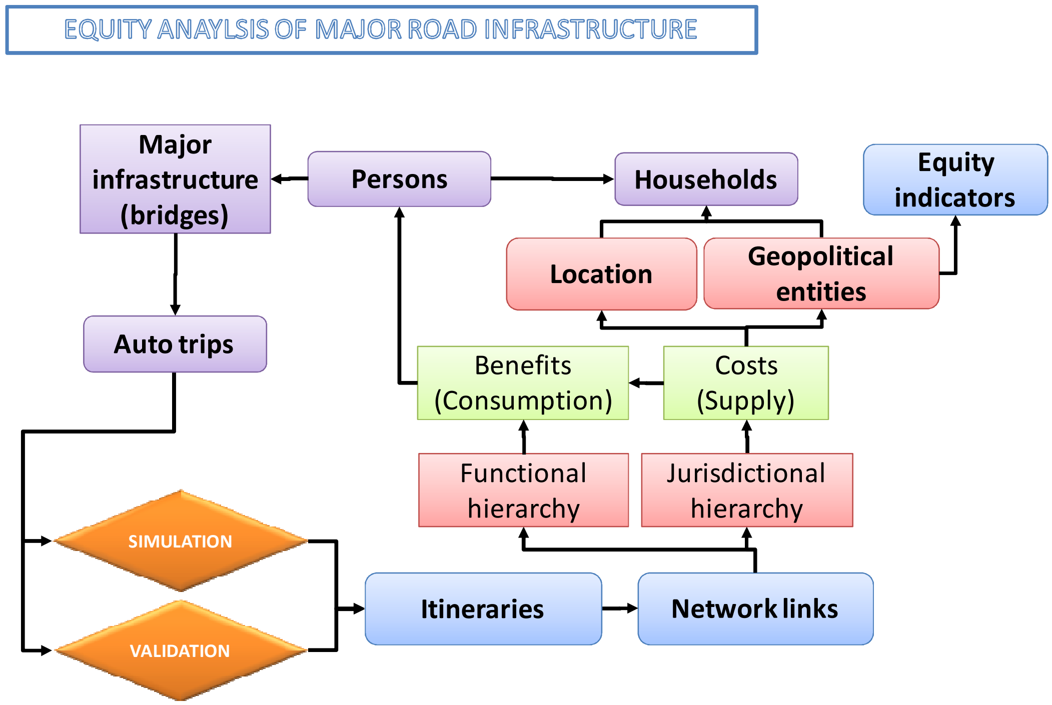


Figure 3.1: Equity analysis of major road infrastructure

3.1 Equity and Transportation Planning

Engineering is primarily concerned with the search for improved efficiency. The power of the technology which engineers are trained to develop and exploit lies in its ability to render previously difficult tasks simple thereby liberating resources for other purposes. The progress of technology is driven by the discovery of more efficient ways of accomplishing important tasks. When it comes to human systems, however, improved efficiency is rarely the only indicator of progress.

One feature that distinguishes a human system from other naturally occurring systems is its desire and ability to exist inefficiently. It is widely accepted that living creatures are the product of a ruthless process of natural selection where the strong become stronger and multiply while the weak become weaker and eventually perish. From a thermodynamic perspective, this system is ideal because the elements which function well are favoured while the defective components are rapidly eliminated. Waste is effectively minimized. Within human communities, by contrast, there exists a sense of empathy or at least a sense of respect toward other members of the same community. The survival of the collective is not always paramount and individuals are believed to have an inherent value. This belief is especially strong in highly developed societies where the existence of the collective does not appear threatened. Respect for the individual has given rise to a formal system of rules (laws) embodying concepts of rights and justice and the notion of justice frequently extends to areas beyond the legal realm. For example, the construction of complex civilizations has resulted in (or has been driven by) the evolution of *systems* designed to accomplish specific collective goals. To the extent that he uses or contributes to them, each member of the society has some relationship with these systems and these relationships can be evaluated using the concept of justice, fairness or, in the socioeconomic domain, equity.

Socioeconomic equity is usually associated with the desire to provide equal access of opportunity to all. The achievement of this objective requires that people of limited financial means are able to consume the same amount of basic services as wealthy individuals. Policies which facilitate this type of equity usually involve a transfer of resources from the rich to the poor. Although it does not form an object of universal consensus, this “Robin Hood” conception of fairness has been advocated by numerous philosophers and economists (Rawls, Galbraith, Keynes, etc.) and its degree of acceptance is demonstrated by the important number of initiatives, especially in the areas of health and education, adopted by governments around the world with the aim of ensuring equality of opportunity. Within the specific context of taxation policy, equal access to opportunity is often discussed using the progressive/regressive dichotomy. A progressive taxation system is one where the charge on an individual depends on the individual’s ability to pay. A regressive system charges all income levels equally, thus ensuring that the charge is most strongly experienced by the taxpayers who are the least well-off. The justification for a progressive policy is that the wealthy are those who have benefitted most from the use of collective resources and should therefore pay a greater proportion of the costs than those who are poorer and have benefitted less.

The degree to which road transport constitutes a “basic service” like health care or education is debatable. While life without a car is unimaginable for many people, it is certainly possible to live a very comfortable existence without the benefit of personal motorized transport. The complexities and nuances contained in this last statement are formidable and the transportation engineer is frequently called upon to clarify them. In order to contribute to this task, the present study is not concerned with socioeconomic equity, but rather with the treatment of the population by the large and complex system that is the public road transport service. The relationship between the population and the passenger transport system is a major determinant of the consumption patterns of energy and land. Moreover, the passenger transportation system is a redistributive mechanism which confers benefits on some people while imposing the associated costs on others. It therefore seems reasonable to wonder whether the system, as it currently exists, is equitable.

The focus of this research is limited to a component of the passenger transport system: the major road infrastructure of a medium-sized urban area. Part of the current controversy surrounding automobile use revolves around the hypothesis that the people who benefit from that activity do not assume a “fair” proportion of the costs. Therefore, any method which seeks to measure equity in a transport system must be able to distinguish between the people who benefit from the system and the people who pay for it. It must also be able to clearly define the costs and the benefits of road transport.

3.2 Costs and Benefits of Road Transport

A cost is defined as a quantity that is given. A benefit is a quantity that is received. Implied by both terms is the notion of an exchange: a quantity of x is given in return for a quantity of y. In the context of a transaction (monetary or otherwise) between individuals or organizations, these explanations seem redundant. In context of automobile travel, however, the picture is considerably less clear. When a driver uses his car to complete a trip, he necessarily pays for the privilege and he presumably obtained some benefit equal or greater to the cost otherwise he would not have bothered. But what are these costs? What are the benefits? Where and when did the transaction occur? And who, apart from the driver, participated in the transaction? This section attempts to provide satisfactory answers to these questions.

The domain of economics has developed a formal vocabulary associated with discussions of benefits and costs. Five adjectives are particularly important: total, average, marginal, internal and external. The total cost (or benefit) is a sum over all consumers of a good. The average is simply the total divided by the number of consumers. “Marginal” refers to one additional consumer at current levels of consumption. Internal and external describe the degree to which the costs and benefits of an activity are shared among consumers and non-consumers. These definitions are important in the discussion that follows.

3.2.1 Definition of costs

For the purposes of the present analysis, the total cost of road transportation is subdivided into two categories: external costs and internal or "out-of-pocket" costs. The external costs include all items which are paid for, at least partially, by people other than the driver. One of the most cited external costs is traffic congestion, otherwise described as the consumption of physical road space by private vehicles. Other examples of external costs include traffic accidents, noise, air pollution, greenhouse gas emissions and degradation of the built infrastructure resulting from usage. These costs are borne directly by the territory in which travel occurs. It is widely accepted that an important component of the external costs is internalized through taxation and fees charged by the government that owns the road. The municipal government funds its road network using a tax on property while the provincial and federal governments tax incomes and consumption. Very often, however, these taxes are used to fund not only transport infrastructure but all government programs. In such cases, the public financing mechanism represents an important economic distortion, the fairness of which can be legitimately questioned.

The internal costs of car travel consist of the purchase price of the vehicle, fuel, maintenance, parking, insurance and licensing fees. These costs are covered by the driver and can be calculated precisely in theory although in practice sufficient data are rarely available to perform the calculation on a large scale. An important non-monetary internal cost is the time spent travelling. In many analyses of urban transportation, travel time is assumed to have some calculable monetary value computed using an assumption about the time value of money. A reasoned examination of the implications of this concept when applied to intra-urban passenger travel, however, reveals numerous fallacies.

Prevalent economic wisdom says, with some empirical support, that two initially isolated communities can mutually benefit from increased interaction with each other. The intensity of this interaction will depend, among other things, on the travel time between the two communities. If the time required to travel from one place to the other is reasonable, then there will be interaction and tangible monetary benefit. In this context, an unreasonably large travel time could be associated with a monetary cost although an unattained benefit is a more correct description. Meanwhile, an individual or firm that earns income by transporting goods or people can also impute a monetary cost to travel time resulting from operator's salaries and lateness charges. In the context of day-to-day personal travel, however, it is much more difficult to assign a dollar value to the time spent travelling because there is no associated monetary transaction. Although some consideration should be given to the theory that tenants trade higher rents for lower transportation costs (Alonso, 1964), workers do not bill their employers for the amount of time they spend commuting each day and retailers do not allow shoppers who spend more time on the road to pay less for their purchases.

At the level of private individuals, time is considered valuable for sole reason that it is a non-renewable resource. Every human being has a finite amount of time at their disposal. This amount varies greatly from person to person and individuals have only an approximate idea of how much is available to them. Fundamentally, people are aware that they will eventually die and their interpretation of this constraint has profound effects on the way they choose to live. On a more mundane level, each person needs to devote a certain amount of time to the satisfaction of biological or societal imperatives such as eating, sleeping, going to school and working. The time remaining is devoted to other pursuits which may or may not require travel. If an economic term to describe travel time must be chosen, perhaps the best one is “opportunity cost”. Time spent travelling is nothing more than time that could be spent doing other things. What value, in terms of money, do all these other things have? No one can say.

Nevertheless, a quantitative analysis requires units of measurement. Although the imputation of a dollar value is beyond the scope of this research, it is fairly easy to justify the use of vehicle-kilometres-travelled (VKT) as a proxy measure of monetary cost. Fuel consumption, vehicle maintenance and infrastructure degradation generated by vehicle use are all proportional to the distance travelled by the vehicle. The non-monetary cost represented by travel time can similarly be represented by vehicle-hours-travelled (VHT).

A peculiar characteristic of the public road transport system is that the transaction between the consumer (the driver) and the service provider (the institution responsible for the network) does not occur at a fixed location. The internal and external costs are extracted from the driver over the course of his trips and are therefore distributed in space. The internal costs of travel by car are paid either by the driver in full, or are shared with other members of the driver’s household. The internal costs of travel therefore accrue to the household of the driver. The external costs of car travel are assumed by all the locations through which the car passes. Some of these costs, such as damage to the road surface, are limited to the road network. Other costs, particularly noise and air pollution, are experienced by anyone located near the employed infrastructure.

3.2.2 Definition of benefits

In general, the internal benefit of travel is derived not from the journey itself but from the activity which motivates the journey. As such, the internal benefit of travel is a consumer surplus – the acquisition of some benefit for a lower cost. Such benefits include employment, shopping and leisure activities. The monetary value of these activities is almost impossible to quantify but, if the standard economic model of man is credible, then the internal benefits an individual obtains by travelling are worth at least the internal costs incurred. In the present analysis, costs are measured in terms of VKT and VHT. It is convenient, therefore, to measure benefits using the same units. As with internal costs, internal benefits are experienced solely by the driver, and possibly by other household members. The internal benefits of travel are therefore assigned to the household.

The measurement of external benefits is more difficult because they cannot be clearly identified. While it might be argued that the activity motivating the trip benefits from the participation of the traveller, it is not the transportation system which confers the benefit. The participant benefits the activity by his presence, not by his travel. Additional external benefits are generated for economic sectors that are peripheral to the transport system itself, such as drive-thru restaurants, but these are considered sufficiently unimportant to be neglected in the present discussion.

3.3 The redistributive effects of a public transport network

The redistributive effects of transport reside in the nature of the transaction associated with the definition of costs and benefits. A single transaction does not have a unique location in space and time and it involves multiple parties, not all of whom are willing participants. The forced implication of individuals and institutions in the operation of a system from which they do not benefit directly is due to the spatial nature of the transaction and the fact that the service provider is almost always a government agency.

The term “public transport” in the title of this section refers not only to mass transit systems but to publicly-owned and operated transport systems in general. Although a road system is considered distinct from a transit system, in the vast majority of cases both systems are controlled by branch of government. Publicly-owned transport networks exist because single-operator systems are easier to plan, manage, and control and because the average benefits produced by physical connections between communities are large enough to justify a collective approach to the construction and maintenance of such systems. The associated monetary costs are shared among the benefiting communities. The cost-sharing is administered by a superior level of government which, in the interests of preserving its territorial integrity, must ensure that all the regions it governs are perceived to be treated fairly. Thus, the high-level (central) government uses tax revenues collected from all its constituencies to finance infrastructure intended to serve particular interests at lower levels. In such a context, the discussion of the benefits and costs of transport becomes highly geopolitical.

It should be recalled that the urban road network has a hierarchical structure based on functionality or jurisdiction. The functional hierarchy has a significant influence on the route choice patterns of drivers and the distribution of road traffic over space and time is the direct result of those individual choices. Therefore, it is the driver's interaction with the functional hierarchy which determines the magnitude of costs and benefits associated with the journey. Meanwhile, it is the jurisdictional hierarchy that dictates the allocation of direct costs among geopolitical entities (governments). Moreover, there is a relationship between the jurisdictional hierarchy and the functional hierarchy of road infrastructure. Urban roads (local, collector, arterial) are inexpensive to build and maintain. These types of facility are somewhat divisible in that they can be built gradually as economic resources become available and as demand evolves. They are almost always the responsibility of the municipal government. High-speed, high-capacity infrastructure requires a large investment and only a government with a broad tax base can afford to take on such projects. In addition, facilities of this type usually occupy land in multiple jurisdictions and so a superior level of government is necessary to oversee construction and operation. Also, these facilities are much less divisible than urban road installations. A freeway must be at least several kilometres long and a bridge cannot become operational in stages. As a result, freeways, highways and major bridges are the responsibility of the provincial and federal governments.

Table 3.1 presents some approximate figures to describe the financing of road infrastructure in the province of Québec. Although municipal governments are responsible for 83% of road kilometres, their investment represents only half of the provincial total. The remaining half comes from the provincial government which is responsible for 17% of the network length when measured in two-lane equivalent kilometres. The cost per kilometre for provincial and municipal roads is estimated for illustrative purposes. The higher per-kilometre cost of provincial infrastructure supports the assertion that provincial infrastructure tends to be of a higher functional class (providing greater speed and capacity) than municipal infrastructure.

Table 3.1: Summary of road infrastructure financing in Québec

TERRITORY	**PROVINCE OF QUEBEC**				
JURISDICTION	TLEK:Two-lane equivalent kilometres (000s)[1]		2008 expenditure ($millions)[2]		*Cost per TLEK ($000s)*
Federal	0	0.0%	24	0.4%	
Provincial	24.5	17.1%	2616	49.0%	*93.8*
Municipal	118.9	82.9%	2699	50.6%	*14.3*
TOTAL	**143.4**	**100.0%**	**5339**	**100.0%**	***24.6***

1. GEOBASE digital road network (www.geobase.ca)
2. (Transport Canada, 2009)

The relationship between the functional class of a road facility and political jurisdiction inevitably leads to economic distortion since first-class infrastructure can be built to serve communities which could never afford it without assistance from the superior authority. In large urbanized regions, these distortions are both inefficient and inequitable since they induce evasive household location patterns. A household can evade the burdens (fiscal and other) of living in a jurisdiction supplying large quantities of transport services by choosing to locate in a municipality which provides only a minimal transport service. This tendency is inefficient because it uses collective resources to subsidize increased consumption and it is inequitable because its costs are borne by territories that enjoy no reciprocal benefit.

To summarize, the fundamental question of equity concerns the government’s treatment of its citizens via the transport system for which it is responsible. Some citizens will receive a disproportionate amount of benefit while others will be forced to pay a disproportionate amount of the costs. Whether a citizen belongs to the first group or the second depends not on his income or his travel behaviour but on where he lives. This realization provides a useful starting point for addressing numerous issues of contemporary interest, particularly: urban sprawl, greenhouse gas emissions, energy consumption and sustainable development. There are usually two schools of thought on how these problems should be addressed. One school says that the solution is to be found in measures which increase efficiency, such as marginal cost pricing. The second approach advocates measures which decrease efficiency by reducing road capacity and providing massive subsidies to public transit. The implementation of either philosophy produces winners and losers. The ability to predict the results of the implementation depends largely on an ability to distinguish between the two.

3.4 Description of a quantitative methodology

The development of a quantitative analysis framework for the assessment of equity requires information based on direct observation. For the present study, information relating to travel demand is derived from the Montreal travel survey. Each auto trip recorded in the survey has attributes of point of origin, point of destination, departure time and bridge employed. This information, though detailed, is partial. To construct a complete trip it is necessary to apply a traffic assignment model. Information describing transport supply is the road network developed at École Polytechnique (the simulation network of the previous chapter). This network, although it includes all the streets in Greater Montreal Area, does not provide a complete description of the regional road infrastructure. Some important elements, such as the number of lanes and the control system, are missing entirely. Also absent is information relating to the schedule of the network, meaning the variation in service levels over time resulting from recurrent traffic congestion. Nevertheless, the network does possess a hierarchical structure based on functional class. According to their functional designations, a speed can be attributed to each link. In conjunction with an easily computed length, it is possible to estimate link travel times which are nominal rather than actual and are independent of travel demand. Finally, the supply-demand interaction is represented by a shortest-path algorithm.

This road network model is considerably simpler than the ones most commonly adopted for traffic planning purposes in that it does not account for congestion effects. The model bears a strong resemblance, however, to the one most commonly used by drivers to plan their trips. Currently, none of the major navigation service providers (Google, Mapquest, etc.) provide time-varying travel times or equilibrium computations when they recommend a shortest path. The present model presumes that drivers follow the recommendations of the information tools at their disposal.

The next three sections describe in detail a method for applying detailed though incomplete representations of supply and demand, in conjunction with a simple traffic assignment algorithm, to a quantitative analysis of equitable transportation. The analysis is based on particular characteristics of its three major components: the hierarchical structure of the network, the non-reciprocal nature of the travel demand for the major bridges, and the execution of parallel simulation and validation models.

3.4.1 The hierarchical structure of the network

The beginning of this chapter described two types of hierarchy in an urban road network: jurisdictional and functional. Three jurisdictions share responsibility for the road network in the Greater Montreal Area: federal, provincial and municipal. Not all jurisdictions are responsible for all functional classes of road, as indicated in Table 3.2. Of the four federal bridges, two are classified as freeways (Champlain and Mercier) and two are classified as arterials (Jacques-Cartier and Victoria). The federal network forms a negligible proportion of the total network length, measured in directional kilometres. The provincial government is responsible for all numbered routes. These include all the freeways in the province as well as their access ramps. The remaining numbered routes qualify as highways in rural areas but in urban settings they are effectively arterials or collectors. Provincial roads account for 12.3% of the regional road network. The municipal network consists of all functional classes except freeways and comprises 87.7% of the regional network. Link speeds for simulation purposes are assessed according to the functional class and this correspondence is also shown in Table 2.3.

The functional composition of the network for each region in the Greater Montreal Area is shown in Table 3.2. The table describes the functional composition of each region's road network in terms of directional kilometres. In all cases, the freeway network comprises less than 5% of the total network. The local road network accounts for between 56% and 69% of each regional total, with Montreal having the largest portion of local roads. Arterial roads form the next largest group and are especially important in the two *couronnes*. The "collector" designation is somewhat arbitrary and does not seem to have been favoured by the creators of the digitized network since it accounts for less than 10% of the total network length. The largest network belongs to the *Couronne sud* which contains nearly 11,000 directed kilometres of roadway.

Laval has the smallest network – just over 3,300 kilometres. It is worth noting that Montreal's network is smaller than those of the two *couronnes*.

Table 3.2: Network composition by functional hierarchy

	FREEWAY	RAMP	ARTERIAL	COLLECTOR	LOCAL	Total km
Montreal	3.5%	2.4%	15.2%	9.7%	69.2%	**8956**
South Shore	4.7%	3.5%	16.4%	13.3%	62.2%	**3907**
Laval	3.1%	2.7%	18.5%	7.9%	67.9%	**3308**
Couronne sud	4.0%	1.3%	29.0%	9.7%	56.0%	**10831**
Couronne nord	4.7%	1.6%	25.1%	9.6%	59.0%	**9112**
ALL	**4.0%**	**2.0%**	**22.3%**	**9.9%**	**61.8%**	**36113**

The importance of the three levels of government in each region of the Greater Montreal Area is shown in Table 3.3 where the weight of each jurisdiction is computed using directional kilometres of roadway. The boundaries of the regions were drawn in such a way that all the federal bridges are found within Montreal. They comprise 0.2% of the entire Montreal network. The provincial government is responsible for 7.3% of Montreal's roads and the municipal governments provide the remaining 92.5%. The role of the provincial government is much more significant in the two *couronnes*, where it provides roughly 15% of road kilometres, more than double the proportion on the island of Montreal.

Table 3.3: Network composition by jurisdictional hierarchy

	FEDERAL	PROVINCIAL	MUNICIPAL	Total km
Montreal	0.2%	7.3%	92.5%	**8956**
South Shore	0.0%	10.0%	90.0%	**3907**
Laval	0.0%	11.1%	88.9%	**3308**
Couronne sud	0.0%	15.5%	84.5%	**10831**
Couronne nord	0.0%	14.7%	85.3%	**9112**
ALL	**0.1%**	**12.3%**	**87.7%**	**36113**

It is worth revisiting briefly the access-mobility dichotomy discussed in section 2.3.3.2. Infrastructure designed solely to provide access benefits only the locations it serves. In an urban environment, this infrastructure almost always lies within the jurisdiction of the local authority which is financed by a property tax. This means that the sole beneficiaries of the access infrastructure are also the sole financiers. In addition, the external costs such as noise and air pollution are also borne entirely by the users, since they live adjacent to the access road. To the extent that this cost-benefit distribution is symmetrical, there are no redistributive effects associated with pure access infrastructure.

Mobility infrastructure, by contrast, benefits locations that lie at a great distance from the facility itself. Locations which are in the immediate vicinity of the infrastructure may or may not benefit but they bear almost all the external costs. Furthermore, mobility infrastructure is most often the property of a national, as opposed to a regional or local, authority. The national government is financed by taxes on the income and consumption of all its residents, some of whom are the users of a particular infrastructure element. Moreover, mobility infrastructure is designed to carry large volumes of traffic at high speeds and therefore generate large quantities of air pollution and noise. As a result, the location and positioning of this infrastructure are not trivial.

3.4.2 Non-reciprocal travel demand

This study uses a subsample of observations from a large-scale travel survey. The subsample was defined as the clientele of major elements of road infrastructure, namely the 15 bridges which provide access to the island of Montreal. As a result, the only mode under investigation is the auto-drive mode. In addition, the time period of the analysis is constrained to the a.m. peak. Table 3.4 shows the breakdown of all the trips in the subsample by origin, destination and trip purpose. A very significant proportion (71.5%) of trips using the major bridges during the a.m. peak period consists of work trips destined to the Island of Montreal. The next largest market segment is trips destined to Montreal for non-work purposes (8.3% of the sample). Work trips originating on the island of Montreal (reverse-commuters) account for 12.7% of total bridge demand at the origin. This demand structure is non-reciprocal in the sense that Montreal hosts far more drivers on its territory than it sends to the other regions. As a result, based on the definition of costs and benefits described earlier, it is likely that the territory of Montreal assumes more costs than its citizens receive in benefits. The extent of this phenomenon will be measured subsequently.

Table 3.4: Morning peak travel demand on major bridges by trip purpose

ORIGINS	Work	School	Return home	Other	ALL	Total Trips
Montreal	12.7%	0.3%	2.4%	1.4%	**16.8%**	**1443**
South Shore	13.5%	0.5%	0.0%	1.6%	**15.6%**	**1341**
Laval	18.9%	0.8%	0.1%	2.8%	**22.5%**	**1934**
Couronne sud	21.3%	1.2%	0.1%	2.2%	**24.8%**	**2128**
Couronne nord	17.8%	0.7%	0.0%	1.7%	**20.2%**	**1737**
DESTINATIONS						
Montreal	71.5%	3.2%	0.2%	8.3%	**83.2%**	**7140**
South Shore	4.0%	0.1%	0.3%	0.3%	**4.7%**	**407**
Laval	4.6%	0.2%	0.7%	0.5%	**5.9%**	**507**
Couronne sud	1.9%	0.0%	0.9%	0.3%	**3.2%**	**272**
Couronne nord	2.1%	0.0%	0.6%	0.3%	**3.0%**	**257**
ALL	**84.3%**	**3.5%**	**2.6%**	**9.6%**	**100.0%**	
% purpose	**7235**	**298**	**222**	**828**		**8583**

3.4.3 Simulation and validation models

The traffic model used for the analysis of equity is “naïve” in the sense that it makes only one hypothesis about its functional components. Supply and demand are both represented by observed data and a simple hypothesis about driver behaviour is employed to represent their interaction. In the absence of information about the spatial and temporal distributions of average traffic speeds, no unbiased representation of congestion effects is possible. It goes without saying that the model would be better if more information were available.

Two simulations are performed. The first is dubbed the validation model since it incorporates the bridge declarations from the travel survey in the driver behaviour hypothesis. The validation model is used to estimate real consumption of road transportation induced by the existence of the major bridges. The second is the simulation model which uses an unconstrained hypothesis about driver behaviour. The measures of consumption obtained from the simulation model are compared to those of the validation model in order to assess the consequences of using a predictive algorithm rather than observed behaviour in the analysis of equity.

For both the simulation and validation models, the driver behaviour hypothesis is represented by an all-or-nothing assignment which assumes that drivers are “optimizing optimists” and therefore always choose the shortest-time path. In other words, congestion effects are not modelled explicitly in the simulation and validation models. Although this omission is certain to generate unrealistic traffic speeds, its influence on the ability of the model to accurately reproduce the choice of bridge remains unclear. In the case of the validation model, congestion effects are accounted for implicitly, since a driver's behaviour reflects his or her perception of reality. For example, even if the real travel time of a trip is much higher than the travel time estimated by the validation model, it must be acknowledged that the trip did in fact occur and that it did in fact incorporate the declared bridge. The advantage of using observed behaviour as the basis for a behavioural model is that other choice variables, of which the modeller may be ignorant, are already accounted for in the observed travel patterns.

Once the trips have been assigned to the network, it is necessary to associate every utilized network link with its users and the territory in which it is located. This process involves a simple post-treatment of the trip plan file generated by TRANSIMS. The resulting database of itineraries is illustrated in Table 3.5. This data structure is totally disaggregate in that it contains only the atomic units of analysis: trips, links and territories. The attributes of each of these objects can be associated with each other using this table. The validation model generated 664,077 trip links for 8,583 trips, representing an average of 77.3 links per trip. The simulation model generated 625 296 trip links or 72.8 links per trip.

Table 3.5: Structure of the itinerary database

Trip ID	**Link**	**Sequence**	**Municipal sector**
8	4421	1	113
8	4423	2	113
…	…	…	…
408824	483914	63	109

The itinerary database facilitates computations of consumption and supply. According to the framework described above, consumption is measured at the household of the trip maker. Supply is measured at the location of each link used in the completion of a trip. Both households and network links are assigned to a territory, specifically one of the 100 municipal sectors defined by the Agence Métropolitaine de Transport as part of the travel survey. The municipal sector in which the traveller resides is an attribute of every trip so no additional analysis is necessary. A GIS algorithm is used to assign a municipal sector to each link. An example of the segmentation of an itinerary among multiple territories is shown in Figure 3.2.

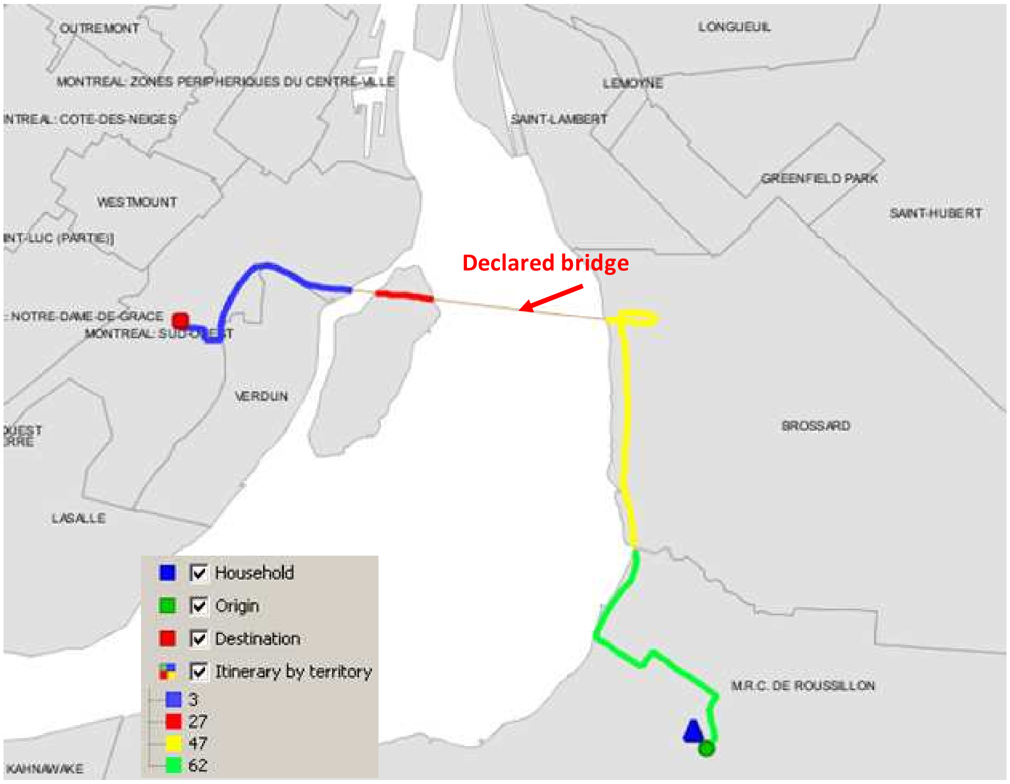


Figure 3.2: The segmentation of an itinerary among multiple territories

In the present analysis, the two quantities of primary interest are the total number of vehicle-kilometres travelled (VKT) and the total number of vehicle-hours travelled (VHT). Both these quantities are distributed among the municipal sectors according to the results of the traffic assignment. A balance sheet of costs and benefits is constructed by counting VHT and VKT generated by the households within a sector as benefits and counting VHT and VKT generated by network links within the sector as costs. A slightly different way of describing this setup is to say that household travel represents observed demand (consumption) while the distribution of this demand over space represents transport supply. An equitable situation is one where a particular territory consumes an amount of transport which is equal to the amount it supplies. Within this analysis framework, total demand (benefits) necessarily equals total supply (costs). Therefore, if there is one territory which consumes more than it supplies, there is necessarily another which supplies more than it consumes. Strictly speaking, such a situation is inequitable since the first territory gains at the direct expense of the second.

3.5 The measurement of distributive effects of road transport infrastructure

The application of the equity analysis methodology consists of four parts. It begins with a thorough characterization of bridge users. This analysis is followed by a detailed description of the consumption patterns of the municipalities and regions which comprise the Greater Montreal Area. Third, the amount of transport supplied by each region to serve bridge users is quantified. Fourth, the distribution of transportation resources over the various functional and jurisdictional networks is examined. Finally, a comparison of the amount of transport supplied versus the amount consumed by each territory generates indicators of equity for bridge-induced road transport.

3.5.1 Characterization of infrastructure users

The identification of the winners and losers resulting from a particular transport policy requires the characterization of the implicated clientele. Each of the major bridges serves a particular transport market. Some important characteristics of this market can be derived directly from the travel survey responses. Ideally the results would be presented by bridge but the fairly large number of bridges (15) and the size of the sample (8,583 trips) means the data are frequently too sparse to represent meaningfully. In consequence, the analysis results are aggregated by screenline and direction. For example, Figure 3.3 shows the demographic profile of bridge users by screenline. There are four screenlines and two directions (inbound and outbound) so the figure describes user profiles for 8 bridge groups rather than 15 individual bridges. The distribution and the inset table show a remarkable symmetry between men and women and a stability of demographic characteristics across screenlines. The data for bridge usage is representative of the general trends with respect to car travel in Greater Montreal. Men constitute the majority of drivers and male drivers are, on average, slightly older than female drivers.

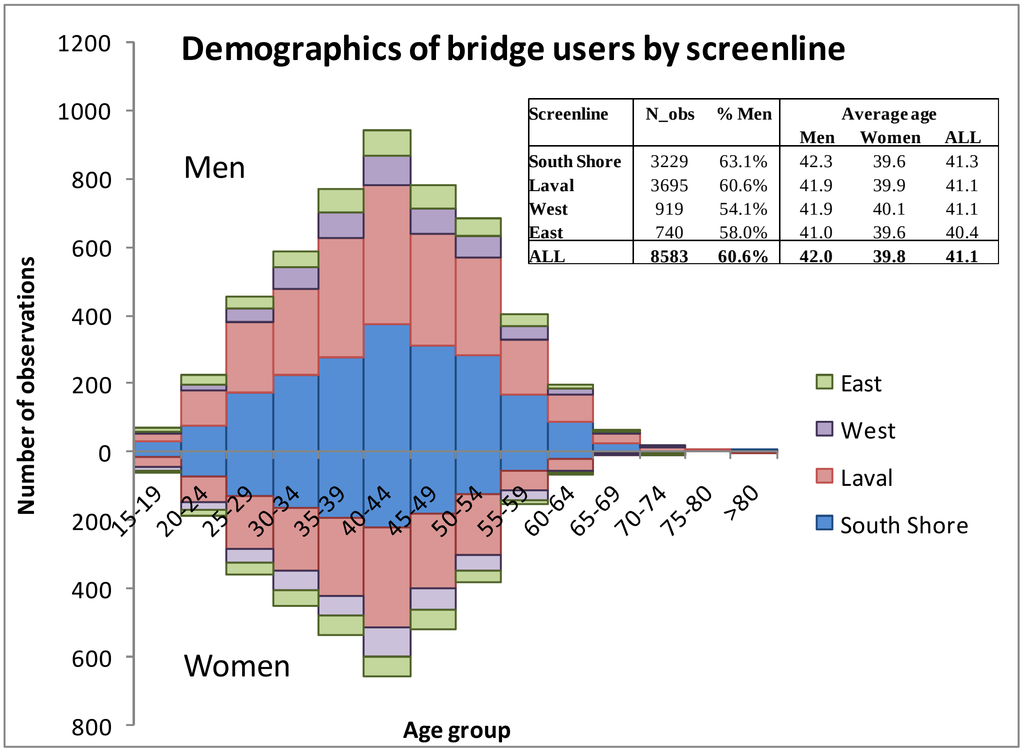


Figure 3.3: Demographics of bridge users by screenline

In addition to personal attributes, the results of the disaggregate traffic assignment model can be used to construct profiles of trip attributes. Figure 3.4 shows the distribution of travel times by screenline and direction as simulated by the validation model. Since congestion is not included in the model, it is certain that the simulated travel times are considerably smaller than the real travel times. Comparisons between bridge groups are nonetheless feasible. Outbound travel times are on average less than inbound travel times but the relative travel time differences between screenlines is similar in both directions. For example, the Laval screenline has the shortest average travel time in either direction and the West screenline has the longest average travel times. Note that the shorter duration of outbound trips cannot be explained by congestion effects. They are due solely to shorter outbound distances.

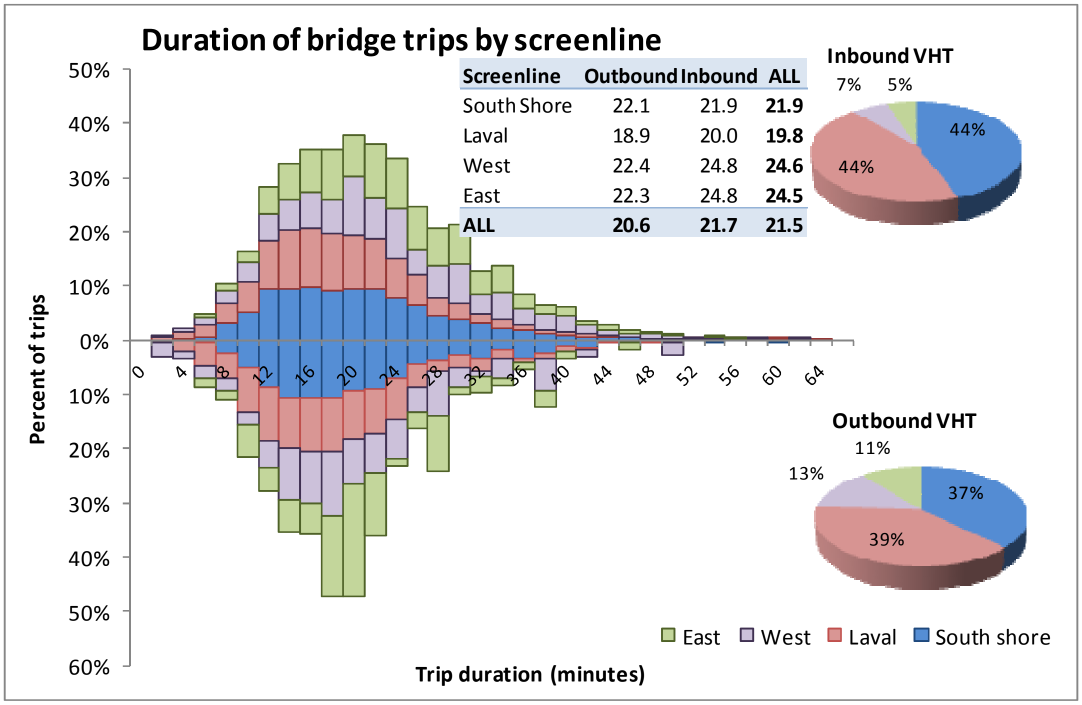


Figure 3.4: Travel times by screenline

The trip length distributions in terms of distance for each screenline are illustrated in Figure 3.5. The distributions for the inbound and outbound directions are fairly symmetrical. The outbound distribution is less smooth due to a smaller number of observations. The inset table shows that Montreal-based trips are, on average, slightly shorter than suburb-based trips especially on the South Shore and Laval screenlines. A possible explanation is the attractiveness of downtown Montreal which has no comparable equivalent in any of the four suburbs. There is a notable difference in bridge usage patterns between regular commuters and reverse commuters. In terms of vehicle-kilometres travelled, reverse commuting occurs mostly on the Laval and South Shore screenlines. The East and West screenlines together account for only 11% of reverse-commute VKT but account for 26% of suburb-based (inbound) VKT. In other words, reverse commuters are most likely to be destined to Laval and the South Shore, a pattern which is likely due to the close proximity of these regions to Montreal.

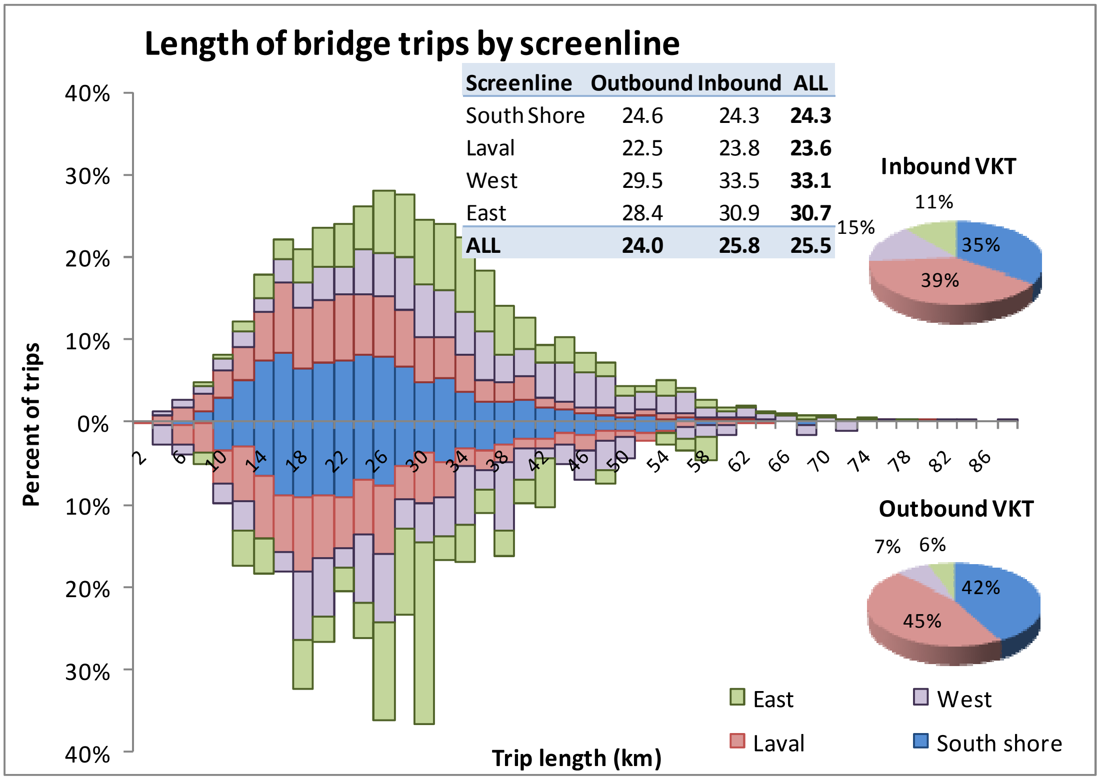


Figure 3.5: Trip distance distributions by screenline

Not all relevant traveller attributes are captured by the travel survey alone. Additional information can be found in the national census (Statistics Canada, 2003). Within the context of discussions concerning road pricing and the value of time, the income of bridge users is of particular interest. Although the survey does ask respondents for the household income, the responses have yet to be validated. The household income of a traveller can be estimated, however, by associating that traveller with a census zone of which the average household income is an attribute.

The aggregation of individual observations into a group (zone) has the effect of replacing a distribution of values with a single mean. Associating individuals with this mean introduces an aggregation bias to the analysis. An attempt is made to minimize this effect by choosing the finest system of census zones for which data are available. These zones, called dissemination areas, usually have populations of less than 2000 and, in urbanized regions they cover a very small geographic area. The Greater Montreal Area contains nearly 6000 dissemination areas. Their small populations and extent limit the intrazonal variation of household income, thereby minimizing the aggregation bias.

Figure 3.6 is a thematic map of average household income using dissemination areas as the spatial analysis unit. Several well-recognized trends are evident. The central and eastern sections of the Island of Montreal are mostly low-income areas. The high income regions on the Island are found primarily on the western half. Large tracts of high income are also found in the peripheral suburbs, especially to the east the St. Lawrence River and directly to the west of Montreal.

Figure 3.6: Average annual household income by dissemination area in 2001

Each bridge user is assigned an estimated household income by associating his household to the dissemination area in which it is located. The average household income of the traveller's household is assumed to be equal to the average household income of the associated dissemination area. Using this method, it is possible to construct income distributions for each bridge and screenline. Figure 3.7 shows the estimated average income of bridge users. Users living in Montreal are plotted separately from users who live off-island. Average income ranges from a minimum value of around $40,000 on the Papineau Bridge (1402) to over $80,000 on the Victoria Bridge (1302). In addition, the estimated income of travellers living off the island is consistently higher than the income of travellers living in Montreal. The trend is supported by Figure 3.8 which indicates that the household income of inbound bridge users is, on average, about $10,000 more than the income of outbound bridge users. The average income of outbound drivers is higher than the average income of inbound drivers on the West screenline only. The discrepancy between inbound and outbound is largest on the East screenline (bridges 1601 and 1602), where the travellers originating in Montreal earn about $12,000 less than their off-island counterparts.

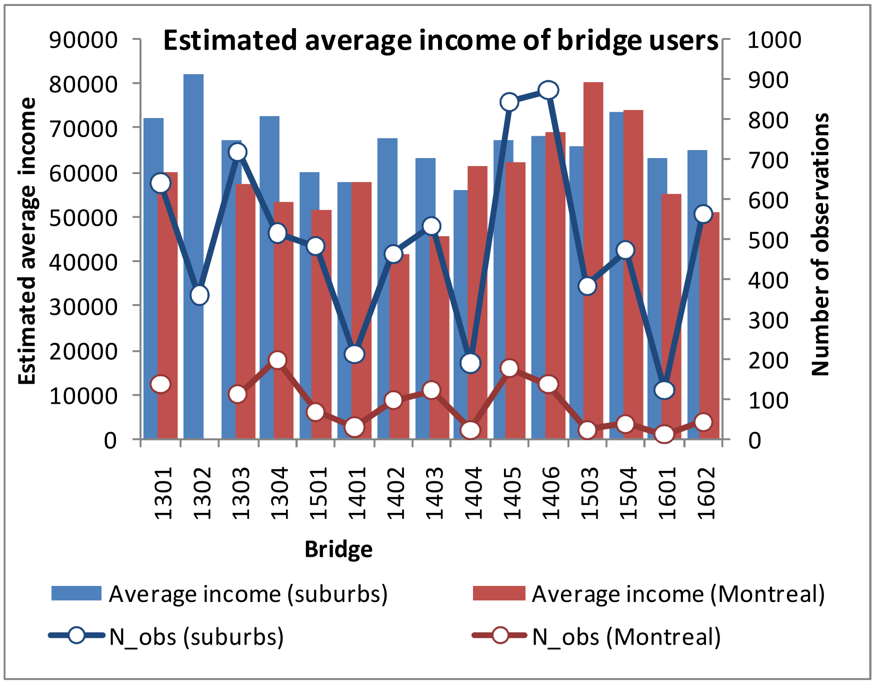


Figure 3.7: Estimated average income of bridge users

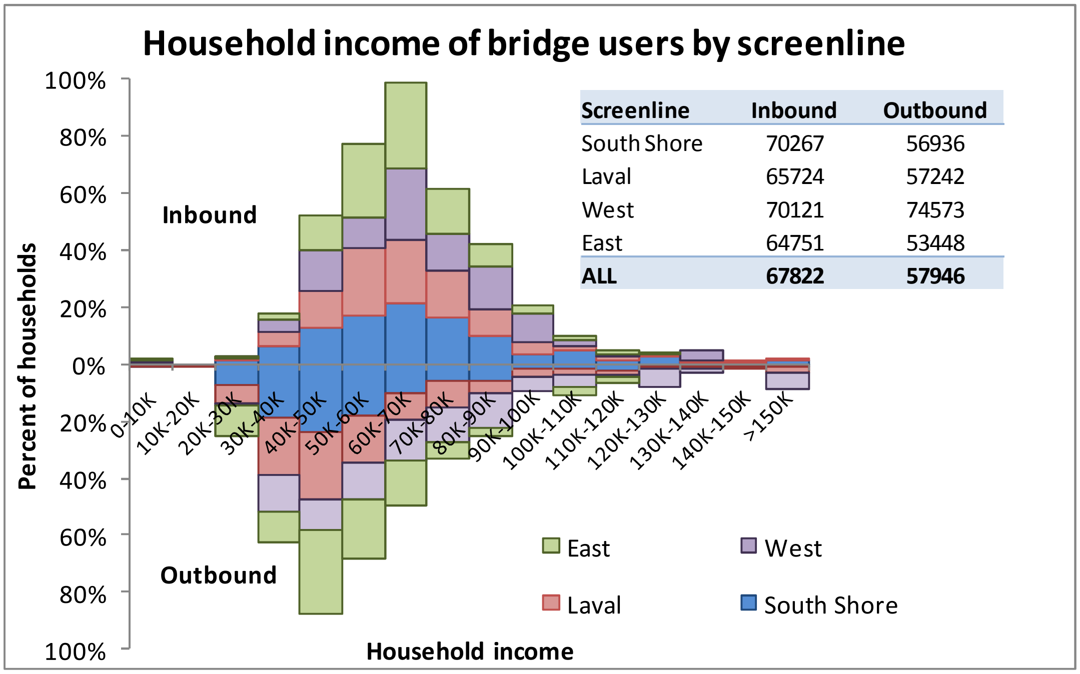


Figure 3.8: Household income of bridge users by screenline

These patterns can be explained to some degree by the income distributions of the five regions, as represented by bridge users (Figure 3.9). The distributions of the two *couronnes* resemble each other closely. Laval's income distribution is comparable to the two *couronnes*. The distribution for the South Shore is considerably flatter but the peak occurs at roughly the same income bracket as the other suburban regions – between \$60,000 and \$70,000. The distribution for Montreal peaks further to the left, indicating a greater incidence of low income. Indeed, the average income of bridge users residing in Montreal is \$7,600 less than the average of the "poorest" suburb (Laval) and \$14,800 less than the average of the "richest" suburb (the South Shore). The lower incomes of Montreal relative to suburbanites contribute to the lower incomes of outbound trip-makers observed previously. The especially high incomes of South Shore residents explain the high average income of users of the Victoria Bridge.

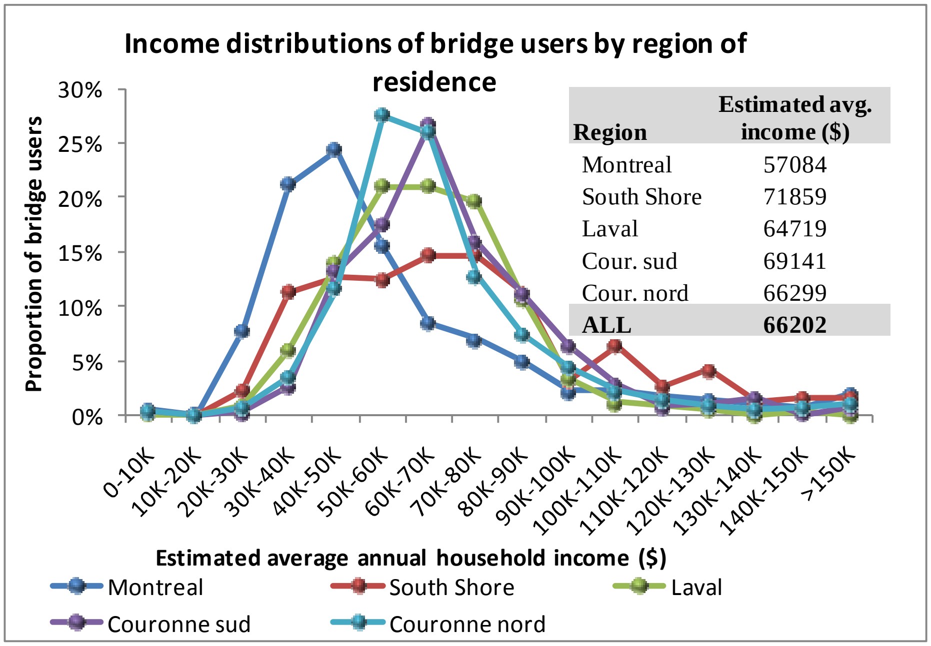


Figure 3.9: Distribution of average income of bridge users by region of residence

Having compiled estimates of bridge user household incomes based on census data, it is worthwhile to compare the results with those obtained using the income declarations in the travel survey. Of the 7,638 households in the validated subsample, 7,061 (92.4%) responded to the question concerning household income. An average value was then computed for each bridge. These averages were compared to those obtained from aggregated census data. The results are shown in Figure 3.10. The figure shows that the household incomes estimated the using the census are consistently lower than those estimated using the travel survey. However, the points for all fifteen bridges lie close to the axis of symmetry. It is not immediately clear which of the two methods gives a more realistic result.

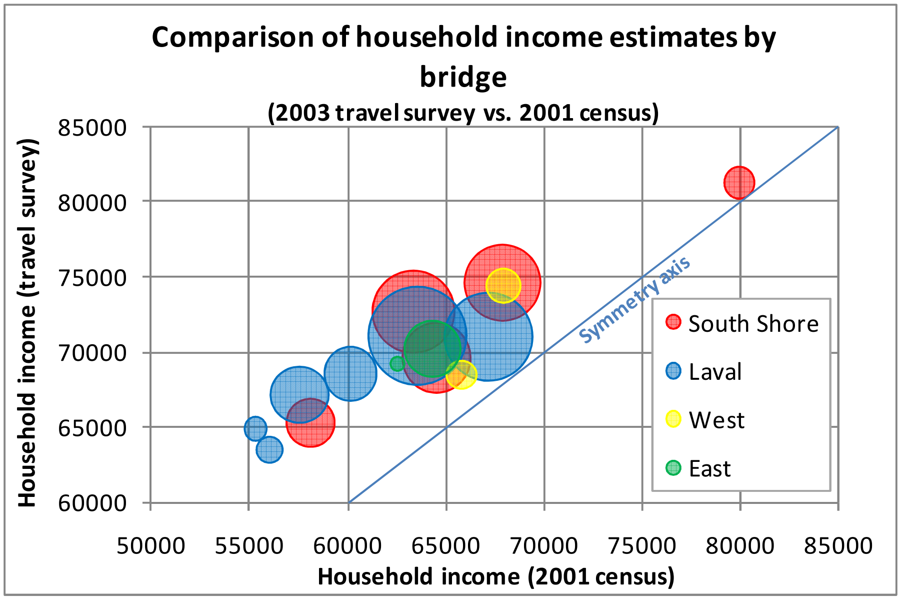


Figure 3.10 : Comparison of household income estimates (survey vs. census) for the users of each bridge

3.5.2 Consumption patterns of municipalities

After constructing detailed profiles of infrastructure users, we proceed to an aggregate analysis of consumption and, in the next section, supply. The units of aggregation are the 100 municipal sectors which make up the Greater Montreal Area and can be further aggregated into five big regions. The consumption of a municipal sector is calculated based on the trips made by residents of the sector. The existence of a money budget and a time budget requires an examination of the question from both the distance and time perspectives. The total consumption is measured in vehicle-kilometres travelled and vehicle-hours travelled. Both these quantities are indicators of the total benefit obtained by the municipality from the existence of Montreal's major bridges. The total benefit per municipality is expressed as the total vehicle-kilometres or vehicles-hours travelled by residents. All these quantities are limited to trips which make of and are made possible by the existence of the 15 major bridges. For this reason, this market of travel demand is subsequently referred to as "bridge-induced" travel.

Figure 3.11 illustrates the spatial distribution of the total benefit induced by the major bridges of Montreal in terms of distance. A similar map (not shown) can be generated using vehicle-hours. The figure shows that the benefits of bridge-induced road transport are much greater in off-island suburban municipalities than on the Island of Montreal. Many of the sectors on the periphery of the region consume less than their immediate neighbours because their interaction with the city of Montreal is not as great. Fundamentally, the figure reflects the fact that morning peak period travel between Montreal and its suburbs (reverse commuting) is much less prevalent than conventional commuting toward the central city.

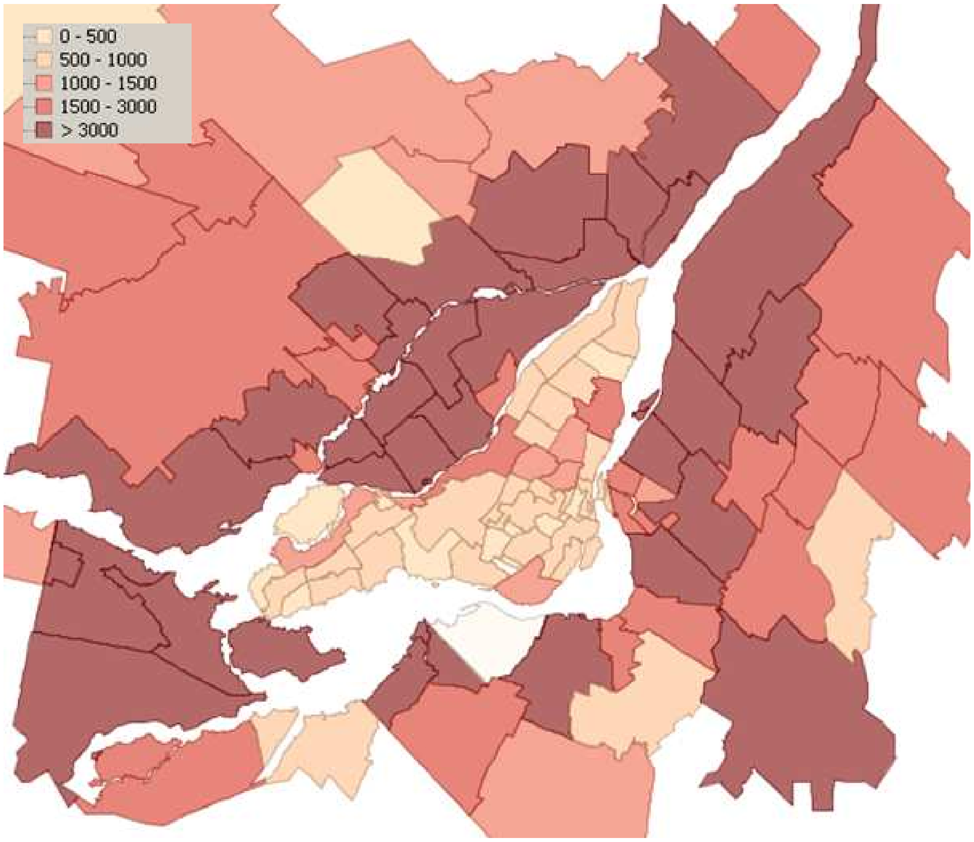


Figure 3.11: Total road consumption in vehicle-kilometres by municipal sector per typical weekday

Figure 3.12 is a breakdown of bridge-induced transport consumption by region of residence and by bridge. Montreal is the only region whose residents make use of all four screenlines. South Shore residents use exclusively the South Shore screenline and Laval residents are mostly confined to the Laval screenline, although a small number makes use of the Pont Charles-de-Gaulle in the East screenline. Most bridges induce significant amounts of travel in three regions but some bridges serve just two. For example, the Victoria Bridge (1302) is the only bridge not used by a single Montrealer during the morning peak period since it provides access into Montreal only. The bridge is used only by residents of the South Shore and the *Couronne sud*.

The inbound Mercier Bridge (1501) serves the *Couronne sud* almost exclusively. The bridges of the East and West screenlines in the inbound direction serve only the *Couronne nord* and *Couronne sud* residents, respectively. The Médéric-Martin (1405) and Louis-Bisson (1406) bridges are notable for the exceptionally large amounts of VKT they induce. Although other major bridges like the Champlain (1301), Charles-de-Gaulle (1602) and Île-aux-Tourtes (1504) also carry important freeways, none induces anywhere near as much consumption. This pattern can be attributed to the bridge users who make very long trips from distant origins in the *couronne nord*.

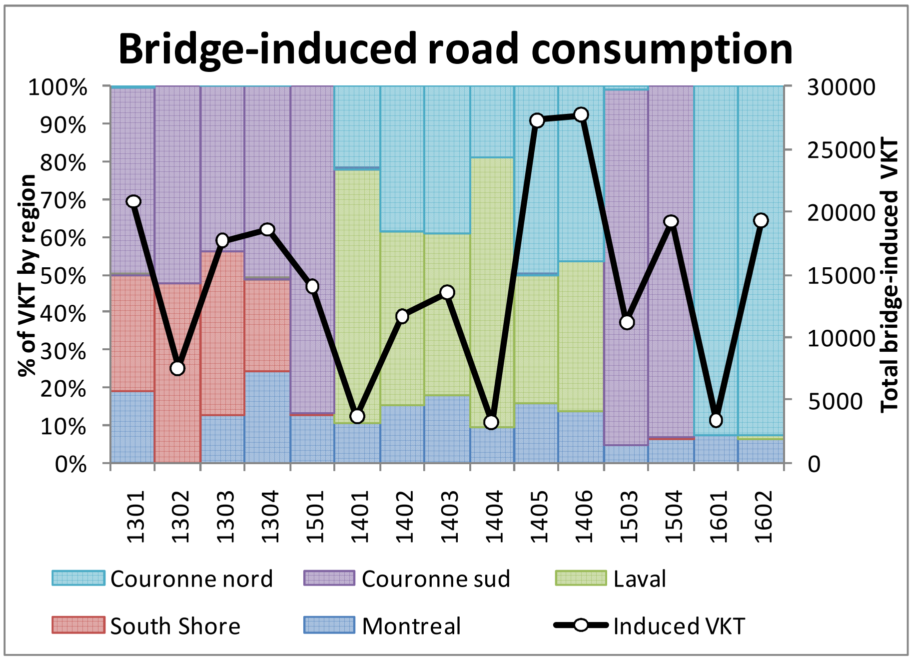


Figure 3.12: Distribution of VKT by bridge and by region

Figure 3.13 shows the consumption patterns by region of residence, measured in vehicle-kilometres travelled. The distributions of VKT for Laval and the South Shore are both compressed to the left as a result of their constrained geographic area and their close proximity to the Island of Montreal. The two *couronnes* are much larger and much further (in network distance) from Montreal and so the distributions are flatter. These two peripheral regions are the primary consumers of bridge-induced transportation. Together they account for nearly half of all the sampled trips and make up 60% of all bridge-induced VKT. The long right tail of the distribution of trips made by Montrealers suggests a diffuse pattern of reverse commuting. Despite the large average length of these trips (23.7 km), the Island of Montreal accounts for only 13% of regional VKT. It is worth noting, however, that Montreal consumes more vehicle-kilometres than the suburban South Shore.

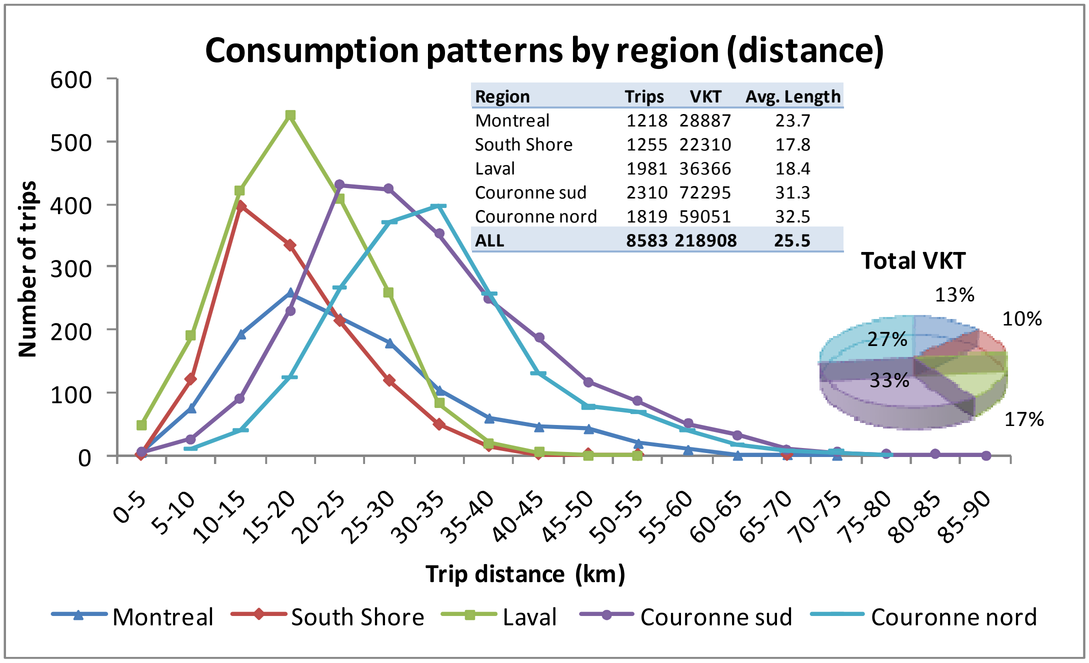


Figure 3.13: Consumption patterns (distance) by region of residence

The same analysis using time as a measure of consumption yields very similar results (Figure 3.14). The only perceptible difference is a slight compression to the left of the distributions for the two *couronnes*, suggesting that drivers living in these regions travel somewhat faster than their counterparts elsewhere. An examination of the average speeds (or levels of service or time costs of travel) experienced by drivers is discussed next.

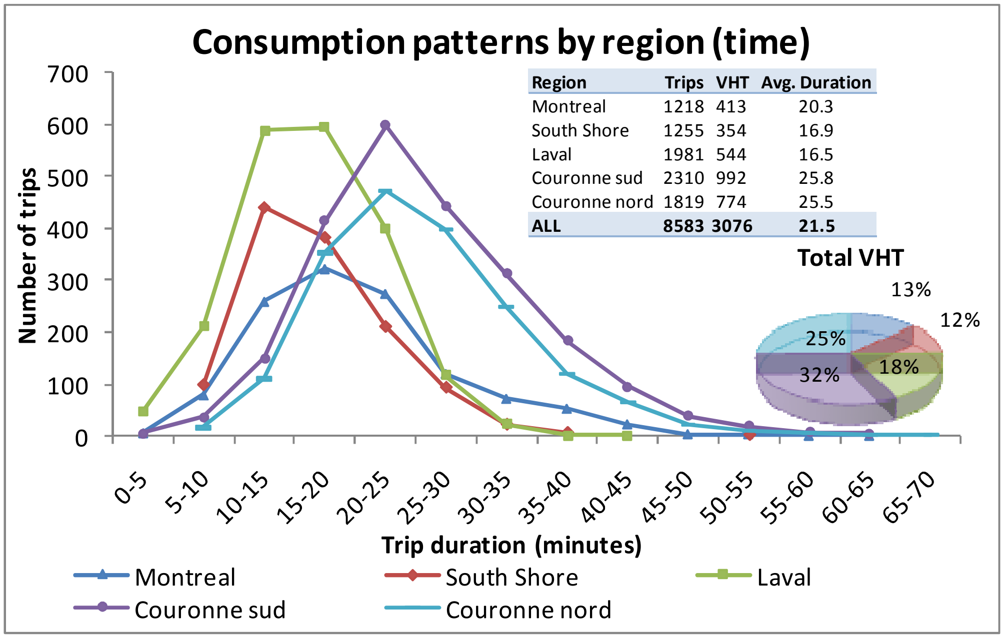
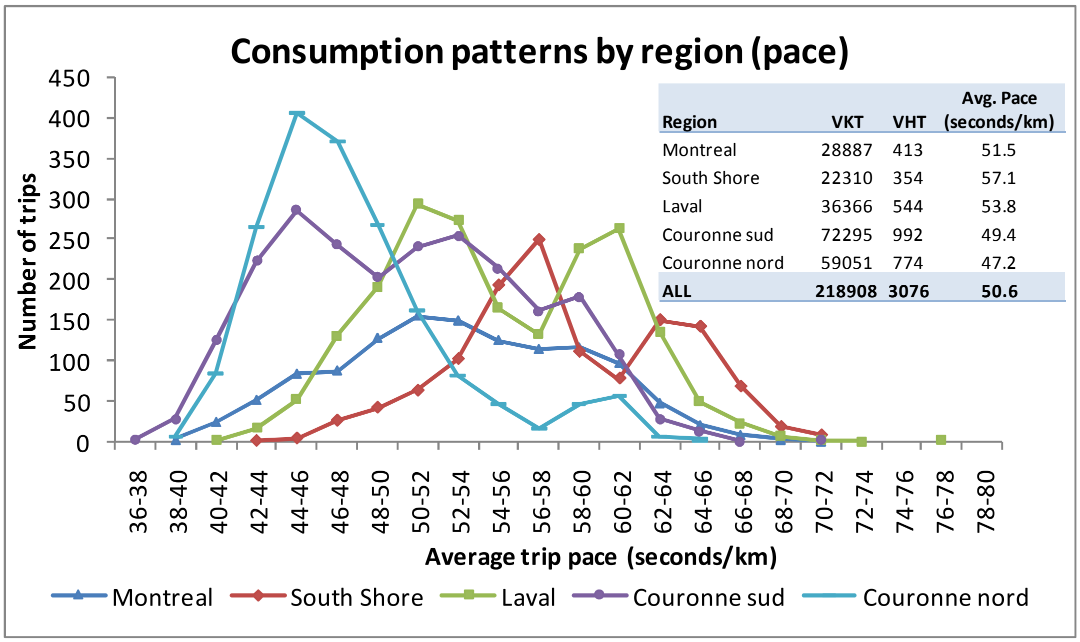


Figure 3.14: Consumption patterns (time) by region of residence

The usual metric used for comparisons of distance travelled and time spent travelling is speed (distance per unit of time) but here we adopt its inverse, pace: the amount of time required to travel one kilometre. The interest of using pace rather than speed is that pace takes the form of a unit price. Generally speaking, modes of transport which are monetarily costly have a low time price. For example, a 500 km flight costs approximately $0.40 and 6 seconds per kilometre travelled. The same trip by car costs about $0.10 and 36 seconds per kilometre. One of the peculiarities of intra-urban road travel is that the time price between different facilities (freeways vs. local roads, for example) varies greatly but the monetary price is essentially fixed. It is for this reason that travellers try to maximize their distance on high-speed facilities, which usually means the same thing as minimizing their travel time. Although an estimation of the monetary cost of travel is beyond the scope of the present analysis, comparisons of the time price (pace) of travel are easily performed.

Another reason for using pace or time price instead of speed is the very precise concept evoke by the term “speed”. Clearly, at many locations the traffic speeds simulated by an all-or-nothing simulation do not resemble the actual traffic speeds experienced during the congested morning peak period. The goal, however, is to examine the characteristics of the offered transport service rather than to evaluate how well the service actually performs. For example, to drivers, a freeway is a multilane uninterrupted flow facility with a posted speed limit of 100 km/h. Traffic congestion may cause the driver to experience speeds below the posted speed limit but the other characteristics of the facility he associates with the posted speed limit remain unchanged. The measures of travel time and time price presented here are meant to illustrate the type of road service that is available to and used by the residents of each region. The absolute values of these metrics are not intended to represent realistic attributes of the traffic stream.

Figure 3.15 shows the distribution of trip pace for bridge users by region of residence. Each region has a distinct distribution. The inset table shows that residents of the South Shore pay the highest time price (53.8 seconds per kilometre) for their use of the major bridges while residents of the *Couronne nord* pay the lowest time price (47.2 seconds per kilometre). The average for Montreal falls between these two extremes.



Region	VKT	VHT	Avg. Pace (seconds/km)
Montreal	28887	413	51.5
South Shore	22310	354	57.1
Laval	36366	544	53.8
Couronne sud	72295	992	49.4
Couronne nord	59051	774	47.2
ALL	**218908**	**3076**	**50.6**

Figure 3.15: Regional consumption measured in terms of pace

Except for Montreal, the distributions for all the regions have multiple peaks which are the result of the functional hierarchy of the road network. Bridges which form an integral part of the freeway system carry trips with a lower average pace (a higher average speed). Trips using bridges which are directly connected to the arterial road network tend to have a higher average pace. This effect is illustrated in Figure 3.16 which shows the pace distribution of the Laval screenline broken down by bridge. The trips with the lowest average pace use almost exclusively the two bridges which are connected to the freeway network at both ends (the Médéric-Martin and the Louis-Bisson). The first peak in the pace distribution includes the four bridges which carry freeways. The second peak incorporates trips that use the Viau and Lachapelle bridges which carry arterial roads. The second peak also contains trips that use the Papineau and Pie-IX but these bridges are only connected to freeways at their Laval ends.

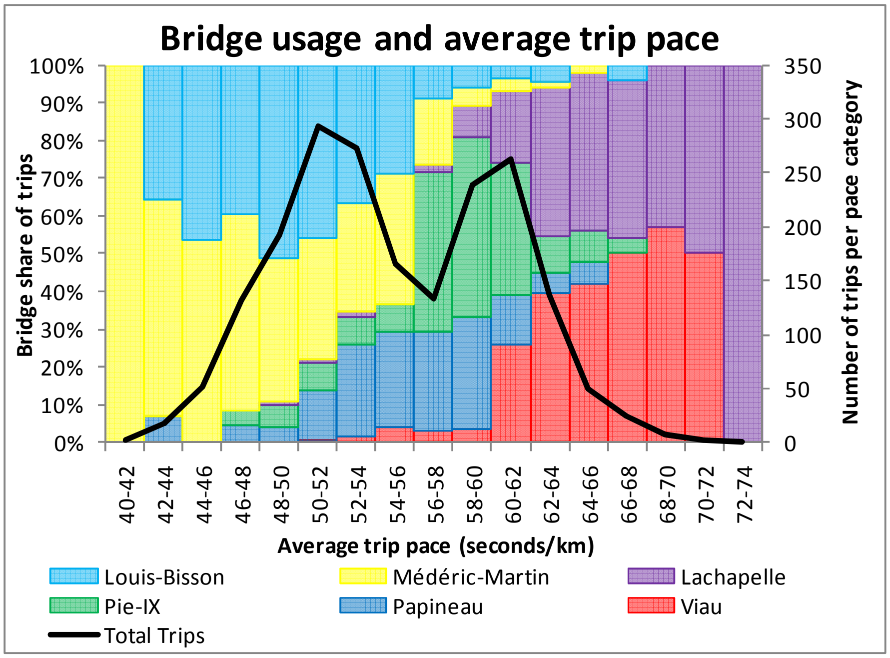


Figure 3.16: The effect of bridge choice on average pace

3.5.3 Transport services supplied by municipalities

The supply of road transport in a municipal sector is quantified using the total vehicle-kilometres or vehicle hours travelled within the boundaries of the sector. This quantity is proportional to the amount of wear on the road system, the amount of air pollution, the amount of noise and the amount of road space devoted to the vehicles of bridge users. Figure 3.17 shows the amount of transportation service supplied by each municipal sector to the clientele of the major bridges. The sectors which are subject to the largest quantities of vehicle-kilometres travelled are located primarily near the bridge heads. Some heavily-burdened zones on the Island of Montreal are not near any bridge but are traversed by multiple major freeways.

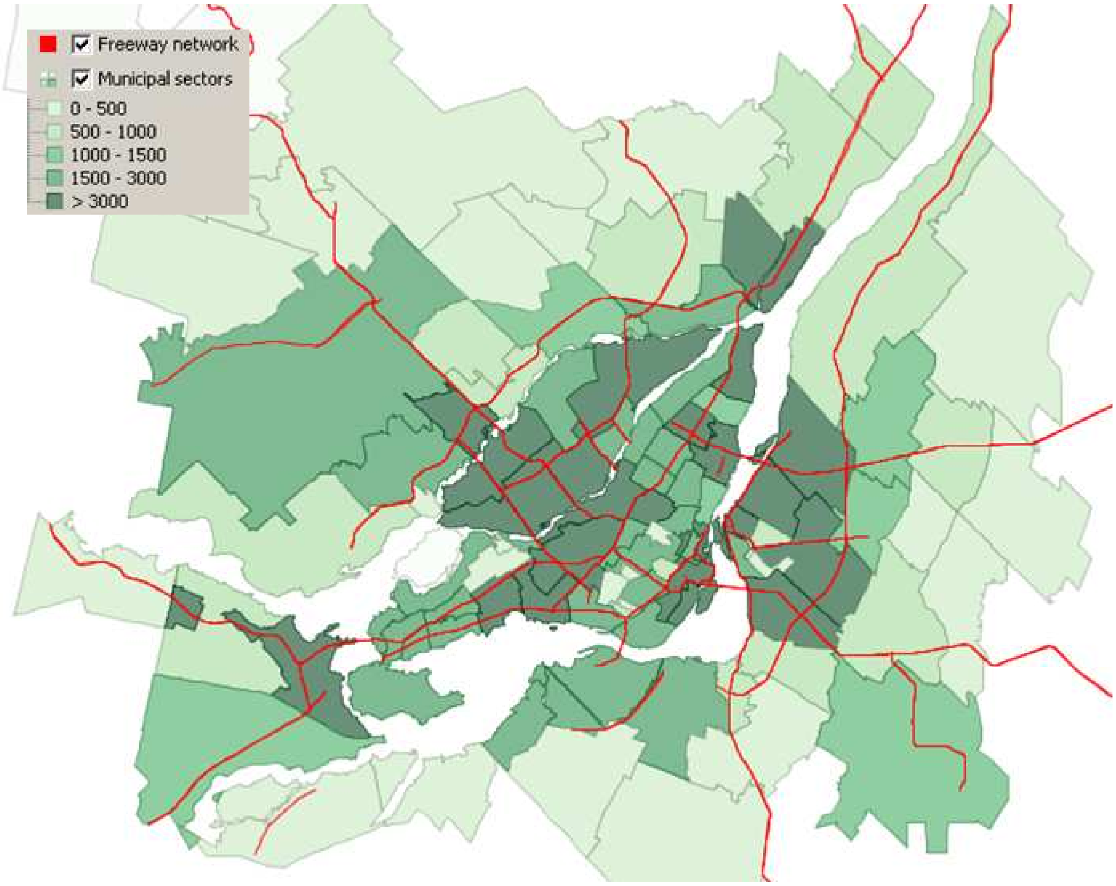


Figure 3.17: The freeway network and the road transport supply (in vehicle-kilometres) by municipal sector

Differences in supply patterns are more obvious when the analysis is performed at the level of the five big regions. Figure 3.18 shows the distribution of path segment lengths per large region in terms of vehicle-kilometres. A path segment is the portion of an itinerary within a single region. Since the subsample of travel demand in this study consists of trips which either start or end on the Island of Montreal, that region contains 8,583 trip segments representing all the trips in the survey sample. Moreover, given the regional geography, it is natural that the Montreal region should account for half of all the vehicle-kilometres supplied to the users of the major bridges. The sharp peak in the distribution of Laval corresponds to the width of the Island of Laval, indicating an important quantity of through trips. Were it not for through-travel, the supply pattern of Laval would closely resemble those of the three other suburban regions. The South Shore is also subject to through-travel, but the phenomenon is not so easily distinguished from the trip length distribution.

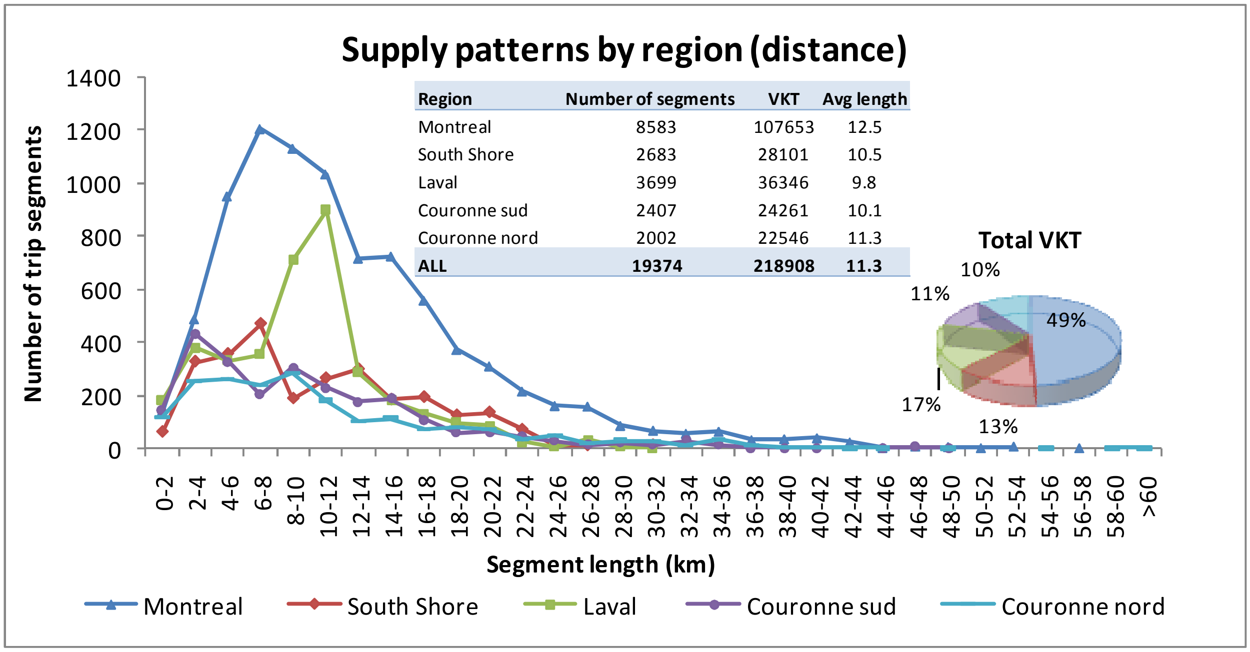


Figure 3.18: Distribution of transport supply (distance) by region

Figure 3.19 shows the regional supply patterns measured in terms of time. Compared to the distance-based distributions of transport supply, the time-based distributions show a certain amount of standardization across regions. All the regional distributions are centred on a 6-8 minute segment length and the distributions have sharper peaks. The regional shares of VHT are almost identical to the regional shares of VKT.

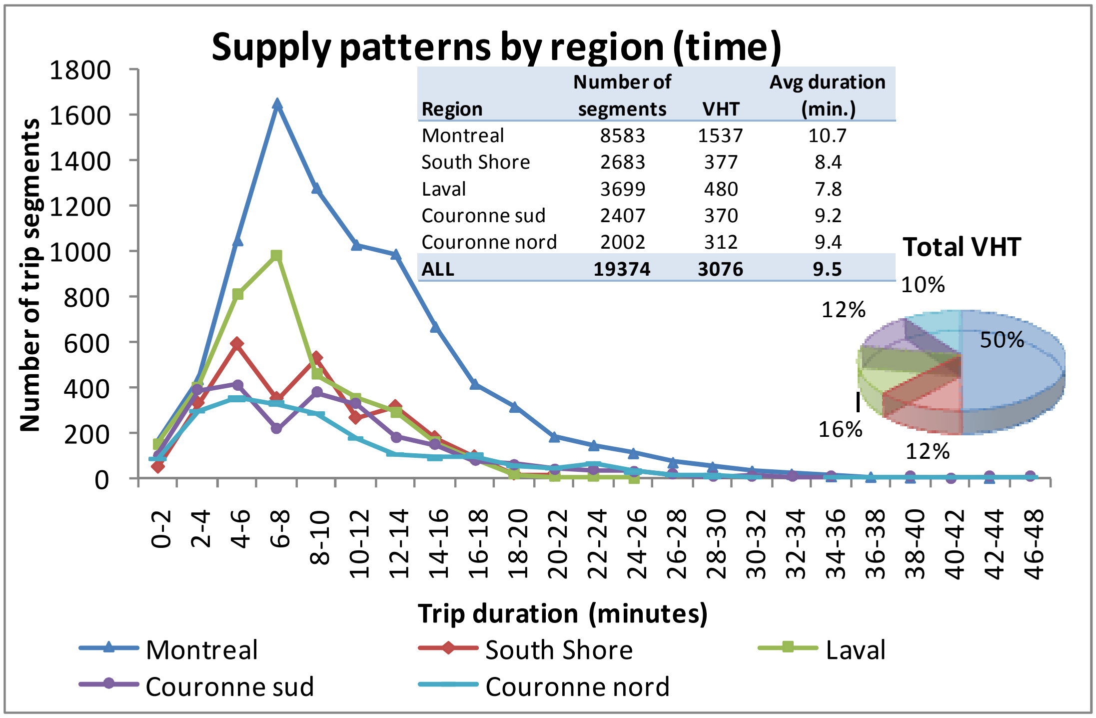


Figure 3.19: Distribution of transport supply (time) by region

Figure 3.20 gives an impression of the time price charged by each region for the use of the infrastructure on its territory. This graphic reflects the functional hierarchy of the network employed by bridge users. For instance, the high concentration of 36 sec/km (100 km/h) trip segments visible in the distribution for Laval is a result of through trips which use only freeways to traverse the region. The peak of the Montreal distribution 60 sec/km (60 km/h) indicates an important quantity of arterial road usage. If pace is considered a price of travel expressed in terms of time rather than money, Laval charges the lowest price (47.5 sec/km), while the *Couronne sud* charges the highest price (54.9 sec/km).

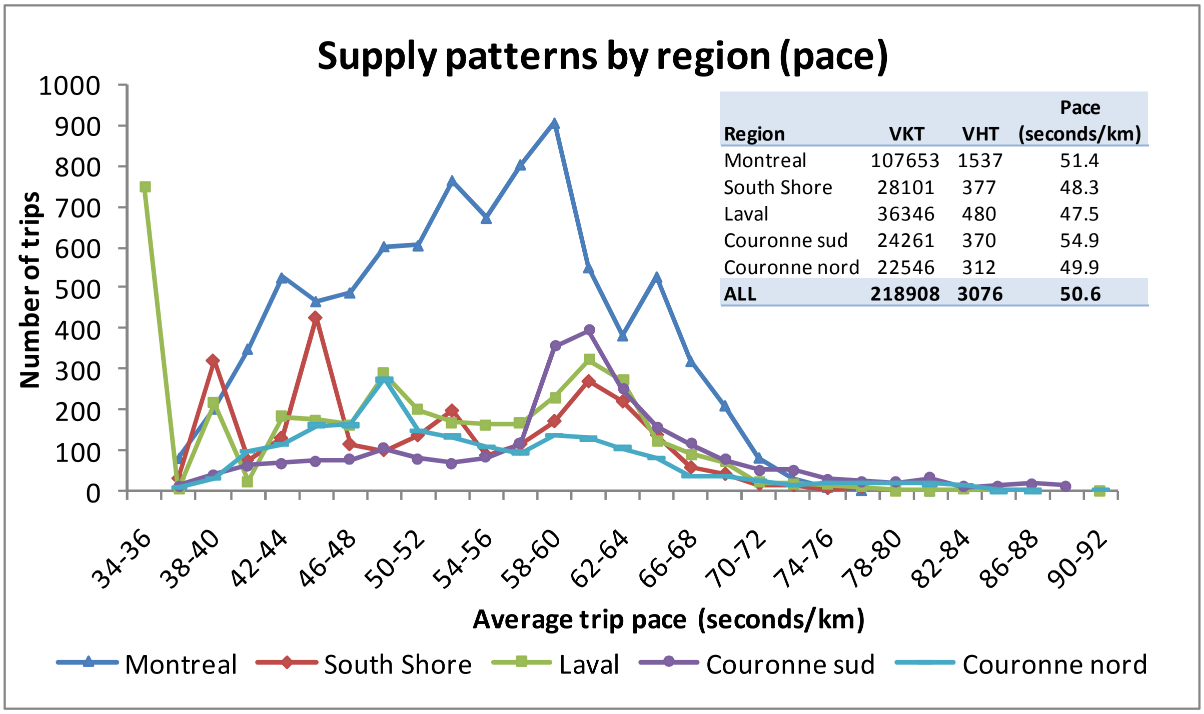


Figure 3.20: Regional supply of road transport measured in terms of pace

3.5.4 The role of networks in the distribution of costs and benefits

The distribution of transportation costs and benefits is a result of the built infrastructure embodied by the road network. Each level in the road network hierarchy contributes differently to the redistribution process and the hierarchical structure is therefore central to the issue of equity. Here, the redistributive effects of the network are examined base on the functional and jurisdictional classifications of its component links.

With respect to road functional class, this research has advanced the intuitive hypothesis that drivers attempt to access the superior network as directly as possible and furthermore, they maximize the distance travelled on the superior network. The trip itinerary database allows for a verification of the hypothesis. Figure 3.21 represents the synthesis of a single average trip from the 664,000 trip segments in the trip itinerary database generated by the validation model. The figure represents a statistical expectation of network usage and shows the distance-weighted probability of using each link type as a function of trip progress. For example, at the half-way point of his trip, there is a 69% probability that a given driver is on the freeway, a 15% probability that he is on an arterial road, a 5% probability that he is on a ramp, a 7% probability that he is on a provincial bridge and a 3% probability that he is on a federal bridge. Another interpretation of the figure is the distribution of VKT by link functional class over the length of an average itinerary. The area of each coloured region represents the total amount of VKT consumed on the corresponding network.

Generally speaking, at the very beginning of trips (within the first 1% of the total length), the local road network is dominant, accounting for 75% of all VKT. The use of local roads decreases rapidly with trip progress. They are not used at all between the first 25% and the last 20% of the average journey. Collector roads follow a pattern similar to local roads while major arterials may be employed during all stages of the journey. Freeways account for the majority of total VKT and are used mostly during the middle stages while the usage rate of ramps stays fairly constant with the proportion of the trip completed. The probability of bridge use is somewhat skewed toward the second half of a trip. This pattern indicates a coherent structure of the model network and, since it is constructed using the observed bridge use pattern, tends to confirm the hypothesis that the distance travelled on the superior network is maximized and that minor roads (local and collector) are mostly used for access purposes at the very beginning and the very end of a trip.

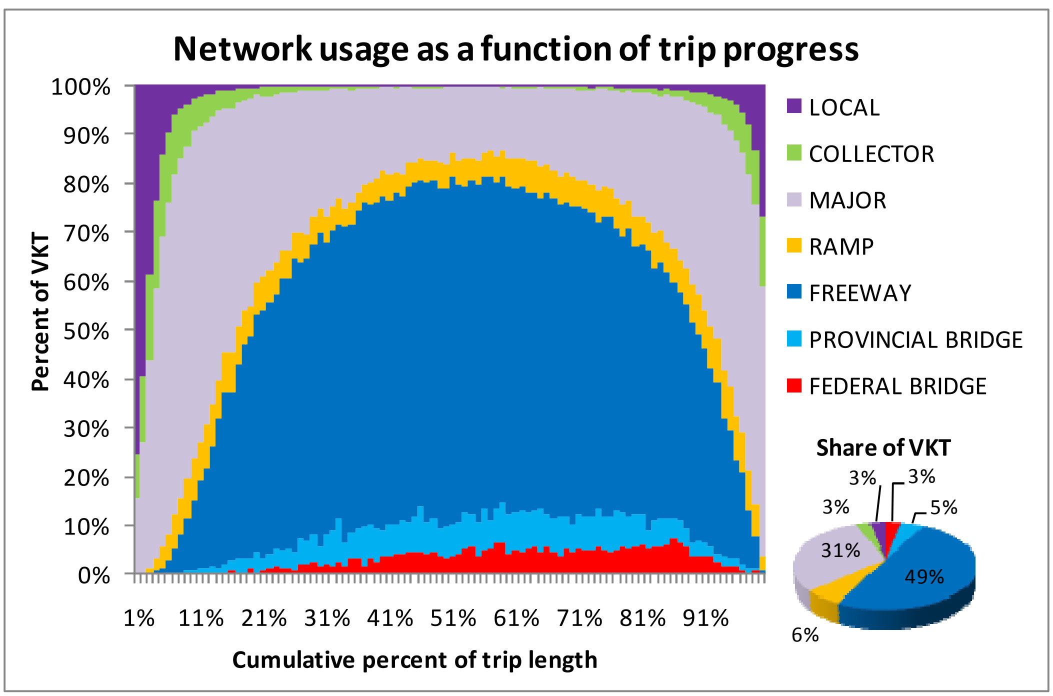


Figure 3.21: Expected network usage as a function of trip progress for an average trip

The redistributive effects can also be analysed using the jurisdictional hierarchy. Figure 3.22 illustrates how different levels of government provide road transport services at different time prices. The figure shows the distribution of consumption in terms of time (above the horizontal axis) and distance (below the horizontal axis) as a function of trip duration percentiles. The shortest trips are grouped into the first percentile while the longest trips are grouped into the 100th percentile. Consumption is measured on the left axis and the average pace of each percentile is measured on the right axis. Several points are worth emphasizing.

First, the use of the municipal network, whether measured in terms of time or distance, does not vary with trip duration. The provincial network, on the other hand, accounts for a progressively larger proportion of consumption as the trip duration and length increase. The higher speed of travel for longer trips is due entirely to the presence of the provincial infrastructure. The federal bridges form a negligible component of total consumption, even though many trips would not be possible if these bridges did not exist. Finally the time price (average pace) of travel decreases steadily as the trip length increases, suggesting the existence of a constant travel time budget. If this budget did not exist, there would be no relationship between consumption and pace.

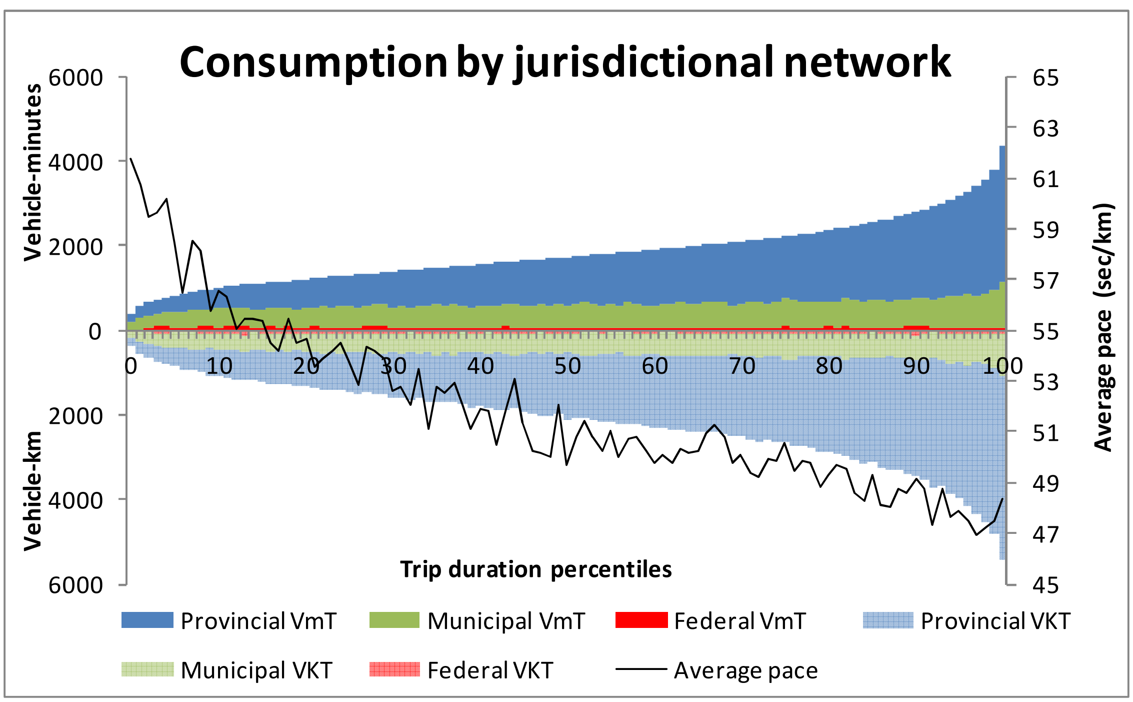


Figure 3.22: Road consumption by jurisdictional network

An examination of the network using jurisdictional class reveals the level of service provided by each jurisdiction as represented by average travel speed or average pace. Each functional class of road has a different free-flow speed representing a time cost of travel and each jurisdictional network is composed of roads of multiple functional classes. As a result of this structure, each jurisdictional network has a unique “supply function”. In traffic assignment models, the most discussed supply function is the one which relates link volume to average speed. In the context of equity, the supply function of interest relates trip distance to average speed since it describes how the price of travel changes with the amount of consumption. Here as well, the analysis is performed using the itinerary database.

Figure 3.23 illustrates the variation of the time price of travel with distance over the four jurisdictional networks. The trip length distributions for each region of residence are also shown. The time price of travel is again represented by pace – the number of seconds required to travel 1 kilometre. Each pace curve represents the variation of the marginal cost of travel as a function of total trip length. The curves are estimated using data grouped by trip length. The average pace for a given network for a particular trip length group is calculated as the time spent divided by the distance travelled on the network. Algebraically,

$$P_{L,M} = \frac{\sum_{i,l_i \in L} t_{i,m \in M}}{\sum_{i,l_i \in L} d_{i,m \in M}} \qquad (2.18)$$

where $P_{L,M}$ is the average pace on network M for a trip length group L, l_i is the length (in km) of trip i, $t_{i,m}$ is the travel time for trip i on network m, and $d_{i,m}$ is the distance covered by trip i on network m.

The figure shows that, on the provincial network, provincial bridges and federal bridges, the cost of travel initially declines sharply with trip distance. For trips longer than 25 km or so, the average pace stabilizes. The large variations in the average pace of the federal bridges for very long trips are due to the small number of observations. For travel on the municipal network, the average pace is almost completely independent of trip length.

The shapes of the pace curves are due entirely to the structure of posted speed limits on the three networks. The federal bridges have lower speed limits than the provincial freeways. In the computation of pace, a bridge has a greater weight in a short trip than in a long trip. Provincial freeways typically account for a smaller proportion of the distance travelled on short trips than the distance travelled on long trips.

The figure illustrates how the marginal cost of travel decreases with distance meaning that, in general, longer trips are made at higher speeds. The fact that the marginal cost function of municipal roads is almost flat suggests that, for long trips, local, arterial and collector roads (under municipal jurisdiction) account for only a small proportion of the total trip distance and, for short trips, the structure of the local network does not provide a decrease in travel cost with distance. In other words, the marginal cost of travel on a major arterial is not much lower than the marginal cost of travel on a local road.

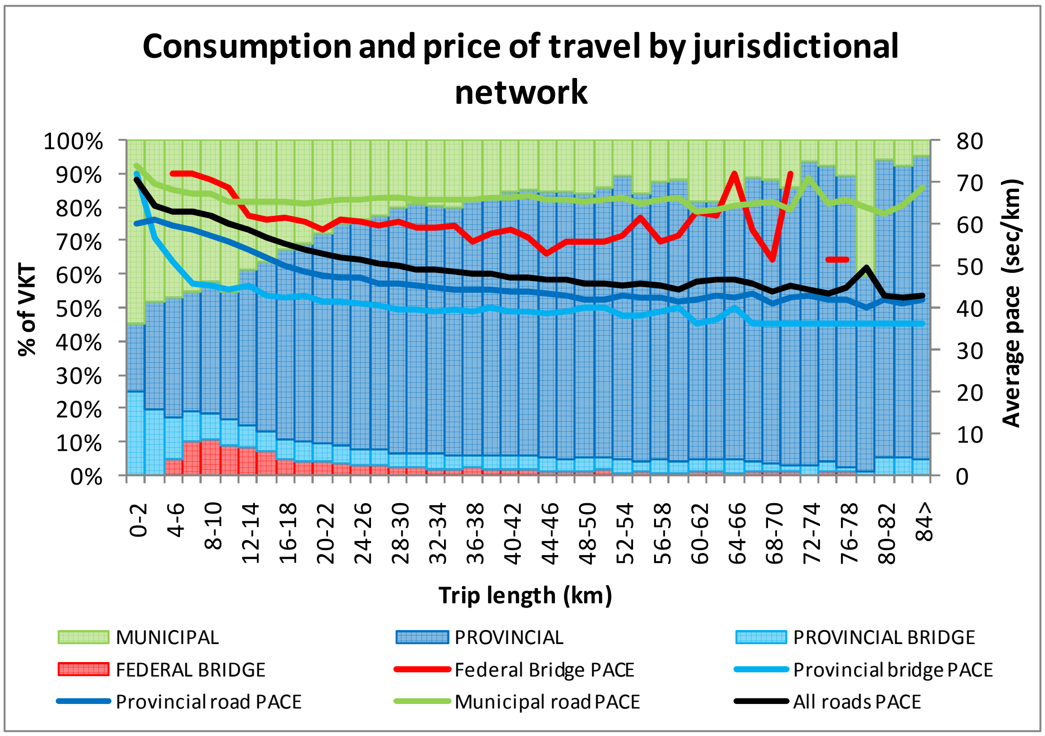


Figure 3.23: Marginal time price of travel for each jurisdictional network

3.5.5 Quantifying equity

Inequity in the territorial distribution of transport-related costs and benefits can be described as an overflow effect. Demand generated in suburban territories flows onto the Island of Montreal and vice versa. Using the method demonstrated by Chapleau & Morency (2004), the nature of these overflow effects is shown in Figure 3.24. The width of each road segment represents the volume of bridge-using vehicles that travelled on the segment during the a.m. peak period, as predicted by the validation assignment model. The volumes on all other links are based on the routes generated by the validation model. Each flow map represents the traffic generated by one of the five regions that make up the Greater Montreal Area and provides an impression of how the external costs of road transport are redistributed in space by the 15 major bridges.

A “transpiration” effect is apparent whereby traffic volumes generated from point sources are channelled first along local roads and then toward high-speed, high-capacity infrastructure. The greatest concentration of vehicles is found on the major bridges. A dispersed use of the municipal/local road network is especially evident in the flow maps of Laval and South Shore residents. The demand from other regions appears more concentrated on freeways and major arteries. While the flow maps for all the suburban regions indicate a concentration of volumes toward central Montreal, the flow map for Montreal residents demonstrates the diffuse nature of reverse commuting. Montreal residents who cross the major bridges during the morning peak period are not channelled toward any important concentration of destinations.

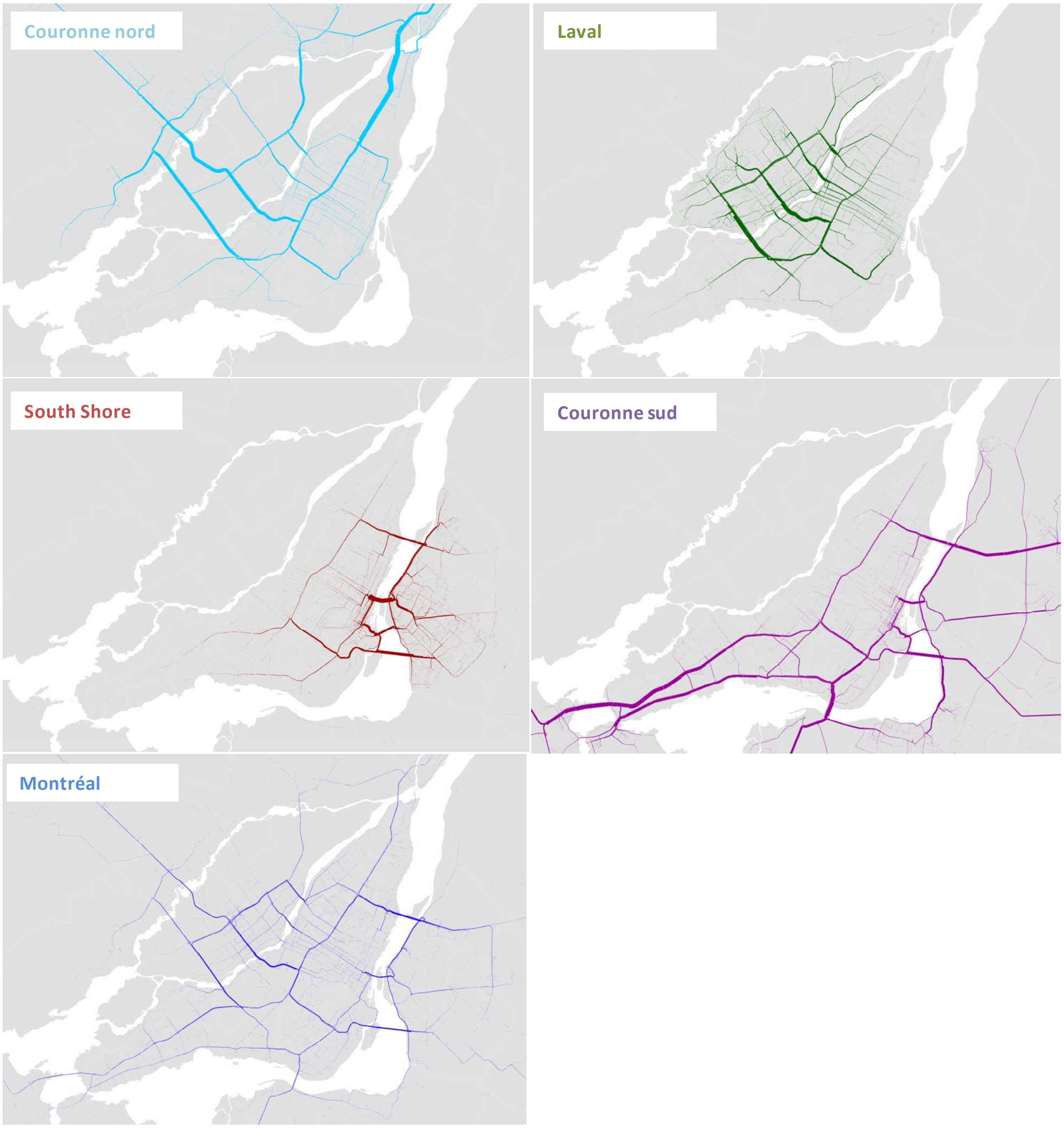


Figure 3.24: Distribution of bridge-induced road consumption over the network for the five major regions of the Greater Montreal Area

An indicator of equity can be constructed by comparing the amount of transport consumed by a jurisdiction to the amount it supplies. One such comparison involves taking the ratio of the two quantities. The result is shown in Figure 3.25. Light blue zones have ratios less than one meaning that they consume less than they supply. Dark blue zones have ratios greater than one and supply more than they consume. Almost all the zones in the latter category are found in the suburban regions while almost all the zones in the former category are found on the Island of Montreal. The three easily distinguished exceptions are the sectors of Outremont, Côte-St-Luc and Dollard-des-Ormeaux.

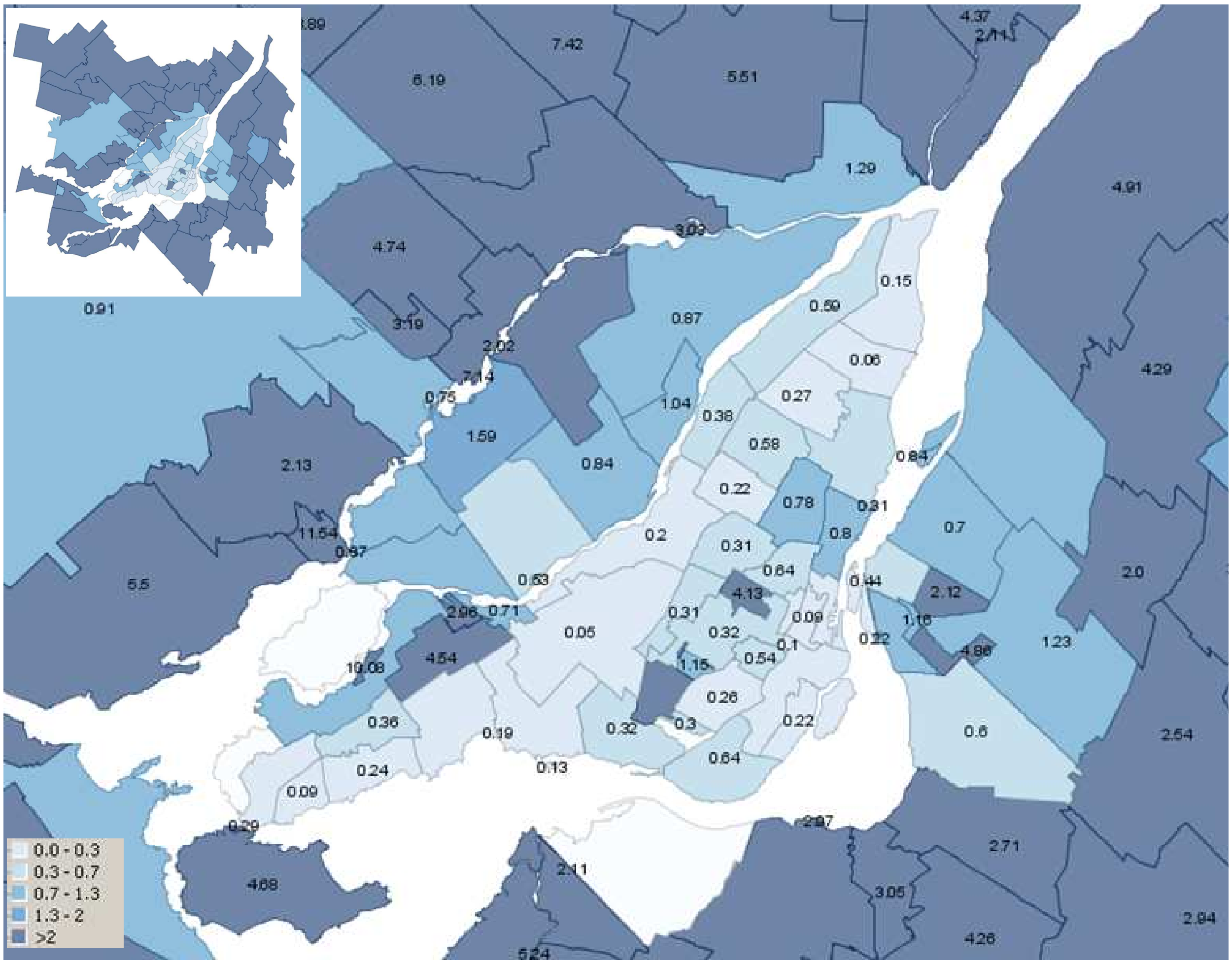


Figure 3.25: Ratio of road transport consumed to road transport supplied

An alternative method of comparison involves plotting the amount of transport consumed against the amount of transport supplied for each municipal sector. Figure 3.26 shows the results of such an exercise. Each point in the figure represents a municipal sector. The consumption of the sector is plotted on the horizontal axis and the amount of transport supplied is on the vertical axis. Points above the axis of symmetry are in surplus since they supply more than they consume. Points below the axis of symmetry are in deficit – they consume more than they supply. Almost all the sectors on the Island of Montreal are in surplus. The three sectors with the greatest surplus are St-Laurent, Ahuntsic on the Island of Montreal and Chomedy in Laval. These three sectors are close to several major bridges and are traversed by multiple major freeways. The sectors within Laval and the South Shore are clustered around the axis of symmetry. The sectors belonging to the *Couronne nord* and the *Couronne sud* are almost all in deficit.

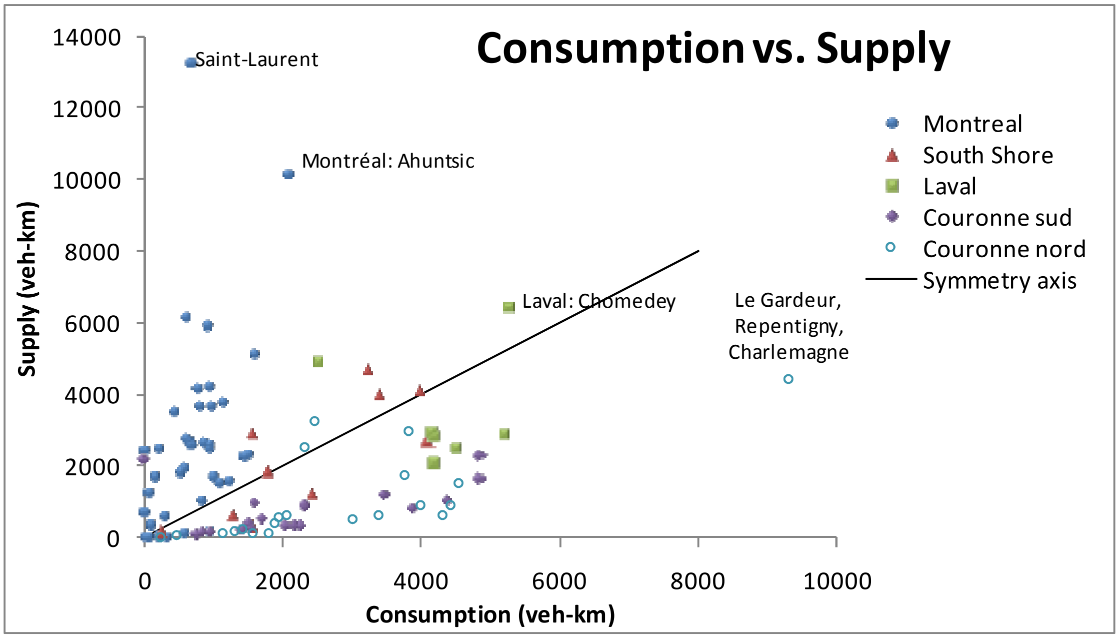


Figure 3.26: Comparison of supply and consumption for the 100 municipal sectors of Greater Montreal

The supply-consumption relationships between each municipal sector can also be examined through the use of a supply-consumption matrix. Figure 3.27 is a visual representation of the supply-demand matrix for bridge users. The vertical axis represents the sectors in which the travellers reside. The horizontal axis represents the sectors within which road transport is consumed. The bars at the top and the extreme right of the figure represent the total consumption or supply for each municipal sector. The diameter of each bubble represents the number of vehicle-kilometres. Each bubble along the diagonal represents travel within the zone of residency. Especially large off-diagonal bubbles represent particularly important imbalances. Some of these bubbles have been numbered in the figure. Bubble 1 represents the large quantity of vehicle-kilometres consumed by residents of *Couronne nord* municipality of Le Gardeur – Repentigny on the Montreal municipality of Pointe-aux-Trembles. This imbalance is a product of the two bridges of the East screenline. Bubble 2 indicates the vehicle-kilometres supplied by

Ahuntsic (Montreal) to the residents of Pont-Viau/Laval-des-Rapides (Laval). Ahuntsic hosts the bridgeheads of three of the six Laval bridges. Similarly, bubble 3 represents the consumption of Chomedy (Laval) residents on the territory of St-Laurent (Montreal). Finally, bubble 4 represents the consumption of Ile-Perrot (Couronne sud) residents on the territory of Pointe-Claire resulting from the use of the Pont Galipeault. In addition to these sub-regional interactions, some broader trends are also apparent. Although Montreal residents consume much less road transport than their suburban counterparts, the Island of Montreal is the major supplier. *Couronne sud* and *Couronne nord* residents consume a significant amount of vehicle-kilometres on the territories of the South Shore and Laval, respectively, indicating an important amount of through traffic on these two latter territories. The large number and size of bubbles along the matrix diagonal indicates that an important portion of consumption is supplied by the municipality of residence.

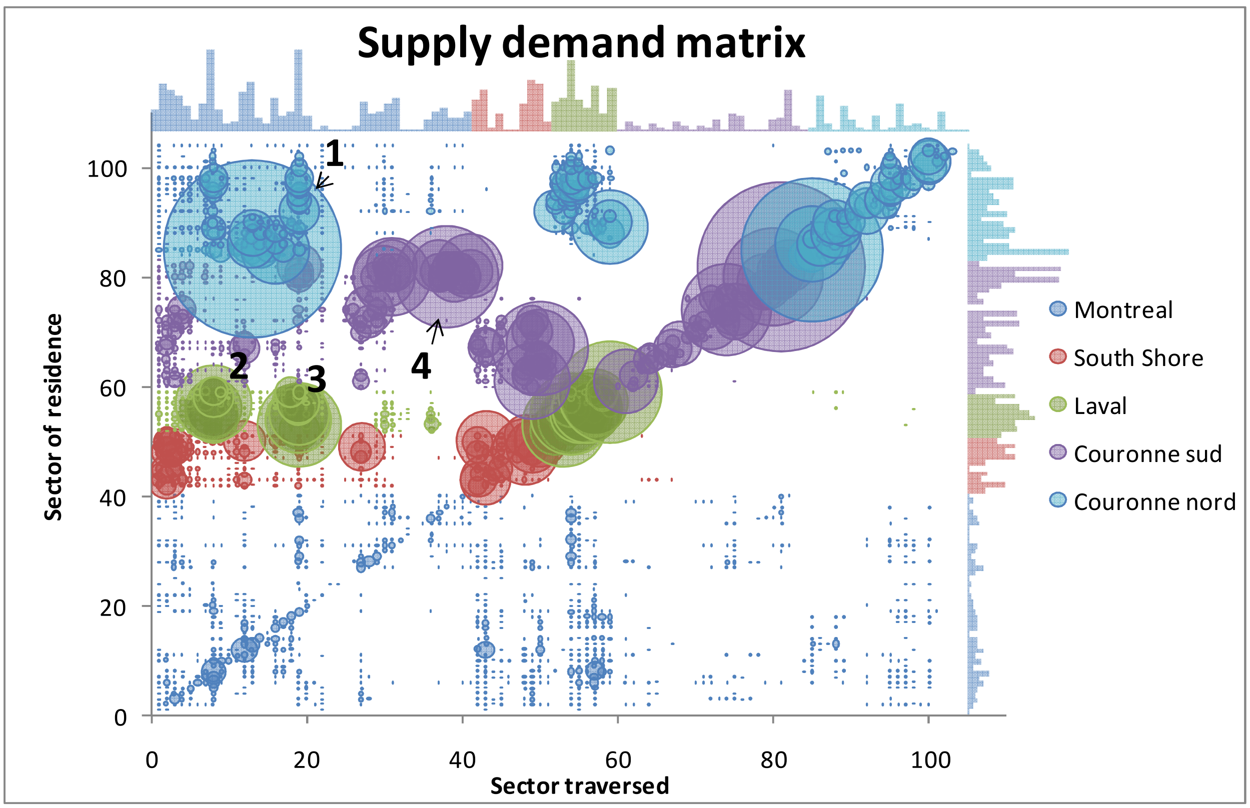


Figure 3.27: Supply-demand matrix of bridge trips

Road transportation consumed by households is supplied by territories via the road network, which is the shared responsibility of multiple jurisdictions belonging to one of the three levels of government. Table 3.6 shows the distribution of consumption over the various jurisdictional networks and regions of Greater Montreal. The leftmost column contains each type of network. The municipal and provincial networks are further subdivided by region. The percentages represent portions of total regional consumption. For example, 5.3% of all the VKT consumed by Montreal-based bridge users is supplied by the provincial network located on the territory of the *couronne sud*.

Globally, the provincial infrastructure carries 74% of the total vehicle-kilometres travelled whereas the municipal network carries 23%. The federal government provides the remaining 3%. The regions of the Greater Montreal Area are each served differently by the three levels of government. For example, the residents Laval and the South Shore consume around twice as much VKT on the provincial network as on the municipal network. By contrast, residents of the two *couronnes* consume four times as much VKT on the provincial network compared to the municipal network. Reverse commuters living in Montreal are also highly dependent on the provincial network. The reliance on the provincial infrastructure is related to the average trip length. The two *couronnes* are most dependent on the provincial network because they are located furthest from Montreal. The South Shore and Laval are the least dependent on the provincial network. The provincial infrastructure located on the Island of Montreal supplies 29.8% of all VKT consumed in the region, the largest share of any single network. Because all trips in the sample begin or end on the Island of Montreal, its municipal network is the only one which contributes significantly to the supply of transportation in the service of the major bridges (11.8% of total VKT).

With respect to the redistribution of external costs using the provincial network, the *couronne sud* displays the least “equitable” tendency. It consumes 34% of its VKT using the provincial infrastructure on the territory of Montreal while the provincial network on its own territory accounts for 22.3% of its consumption. The other three suburbs consume almost equal amounts on their own territories and in Montreal (around 30% of regional consumption in each case).

The supply-consumption patterns for the municipal network show that those of Laval and the Couronne nord are most inequitable since they each consume more VKT on Montreal’s municipal network than on their own. The other two suburbs use their native municipal network more than Montreal’s.

Table 3.6: Distribution of VKT consumed by jurisdictional network and region of residence

SUPPLY BY TERRITORY AND JURISDICTIONAL NETWORK (VKT)	CONSUMPTION BY REGION (VKT)						
	Montreal	South Shore	Laval	Couronne sud	Couronne nord	TOTAL VKT	
FEDERAL BRIDGE	**3.3%**	**13.2%**	**0.0%**	**3.9%**	**0.0%**	**3.2%**	**6997**
PROVINCIAL BRIDGE	**5.6%**	**3.4%**	**5.8%**	**4.4%**	**4.6%**	**4.7%**	**10365**
PROVINCIAL	**65.9%**	**51.7%**	**60.7%**	**73.5%**	**77.9%**	**69.0%**	**151114**
Montreal	*28.0%*	*24.7%*	*30.3%*	*34.0%*	*27.5%*	***29.8%***	***65130***
South Shore	*10.9%*	*26.8%*	*0.0%*	*17.2%*	*0.0%*	***10.0%***	***21831***
Laval	*15.0%*	*0.0%*	*30.0%*	*0.0%*	*23.2%*	***13.2%***	***28977***
Couronne sud	*5.3%*	*0.3%*	*0.0%*	*22.3%*	*0.0%*	***7.8%***	***17117***
Couronne nord	*6.6%*	*0.0%*	*0.3%*	*0.0%*	*27.1%*	***8.2%***	***18059***
MUNICIPAL	**25.3%**	**31.7%**	**33.5%**	**18.1%**	**17.6%**	**23.0%**	**50432**
Montreal	*15.1%*	*15.6%*	*17.6%*	*7.9%*	*9.5%*	***11.8%***	***25791***
South Shore	*3.7%*	*16.0%*	*0.0%*	*0.8%*	*0.0%*	***2.6%***	***5641***
Laval	*3.1%*	*0.0%*	*15.9%*	*0.0%*	*1.2%*	***3.4%***	***7369***
Couronne sud	*1.9%*	*0.1%*	*0.0%*	*9.5%*	*0.0%*	***3.3%***	***7144***
Couronne nord	*1.4%*	*0.0%*	*0.0%*	*0.0%*	*6.9%*	***2.0%***	***4487***
TOTAL VKT	**28887**	**25107**	**36366**	**69497**	**59051**	**218908**	
	13.2%	**11.5%**	**16.6%**	**31.7%**	**27.0%**		

When consumption is measured in terms of VHT (Table 3.7) rather than VKT the picture changes somewhat. The provincial share of supply drops to 34% because of the higher speeds on freeways. The municipal share of supply increases to 63% because link speeds on the municipal network are lower than on the provincial network. The territorial distributions of VHT consumption are similar to the territorial distributions of VKT consumption. All the off-island territories use the provincial infrastructure within the territory of Montreal more than the provincial infrastructure on their own territories. Consumption on the municipal network is more balanced with each suburban region consuming equal amounts on its own territory and on the island of Montreal.

Table 3.7: Distribution of VHT consumed by jurisdictional network and region of residence

SUPPLY BY TERRITORY AND JURISDICTIONAL NETWORK (VHT)	CONSUMPTION BY REGION (VHT)						
	Montreal	South Shore	Laval	Couronne sud	Couronne nord	TOTAL VHT	
FEDERAL BRIDGE	**3.7%**	**14.9%**	**0.0%**	**4.6%**	**0.0%**	**3.8%**	**118.1**
PROVINCIAL BRIDGE	**4.7%**	**3.1%**	**4.3%**	**3.7%**	**3.6%**	**3.8%**	**118.0**
PROVINCIAL	**59.6%**	**45.6%**	**55.2%**	**67.3%**	**71.7%**	**29.9%**	**918.8**
Montreal	*26.1%*	*23.1%*	*28.5%*	*30.3%*	*26.7%*	***15.0%***	***461.3***
South Shore	*9.4%*	*22.2%*	*0.0%*	*14.5%*	*0.0%*	***3.3%***	***102.8***
Laval	*13.2%*	*0.0%*	*26.5%*	*0.0%*	*19.1%*	***4.3%***	***133.7***
Couronne sud	*5.3%*	*0.3%*	*0.0%*	*22.4%*	*0.0%*	***4.4%***	***133.9***
Couronne nord	*5.6%*	*0.0%*	*0.3%*	*0.0%*	*25.9%*	***2.8%***	***87.0***
MUNICIPAL	**32.0%**	**36.4%**	**40.5%**	**24.4%**	**24.7%**	**62.5%**	**1921.4**
Montreal	*19.2%*	*17.6%*	*20.9%*	*10.4%*	*13.0%*	***27.6%***	***848.5***
South Shore	*4.7%*	*18.7%*	*0.0%*	*1.0%*	*0.0%*	***8.6%***	***265.1***
Laval	*3.8%*	*0.0%*	*19.6%*	*0.0%*	*1.5%*	***11.3%***	***346.2***
Couronne sud	*2.3%*	*0.1%*	*0.0%*	*13.0%*	*0.0%*	***7.7%***	***236.4***
Couronne nord	*1.9%*	*0.0%*	*0.0%*	*0.0%*	*10.2%*	***7.3%***	***225.2***
TOTAL VHT	**413.1**	**394.2**	**543.7**	**950.9**	**774.4**	**3076**	
	13.4%	12.8%	17.7%	30.9%	25.2%		

The distribution of costs and benefits among the five regions of the Greater Montreal Area is summarized by comparing the amount of transportation consumed with the amount of transportation supplied in each region. Figure 3.28 is the result of such a comparison in terms of vehicle-kilometres travelled. The bars above the horizontal axis represent the consumption of each region, while the bars below the horizontal axis represent the supply. The pie-chart in the lower right corner shows the distribution of VKT over the four jurisdictional networks. The figure demonstrates that Montreal supplies much more than it consumes, that Laval and the South Shore consume amounts roughly equal to those which they supply, while the two *couronnes* consume much more than they supply. This pattern is consistent with the location of the major bridges within the regional geography. Since all trips in the demand subsample originate or terminate on the Island of Montreal, it is natural that Montreal should supply the greatest amount of road transport resources. It is equally logical, although less obvious, that the regions of the South Shore and Laval should have balanced supply and demand because these two regions provide road infrastructure to residents of the *couronne sud* and the *couronne nord*, respectively. The two *couronnes*, located at the extremity of the Greater Montreal Area, supply road infrastructure to only a small number of reverse-commuters from Montreal. An important segment of the populations of these regions work in Montreal and, as a result, the amount of consumption in the two *couronnes* greatly exceeds the amount of supply.

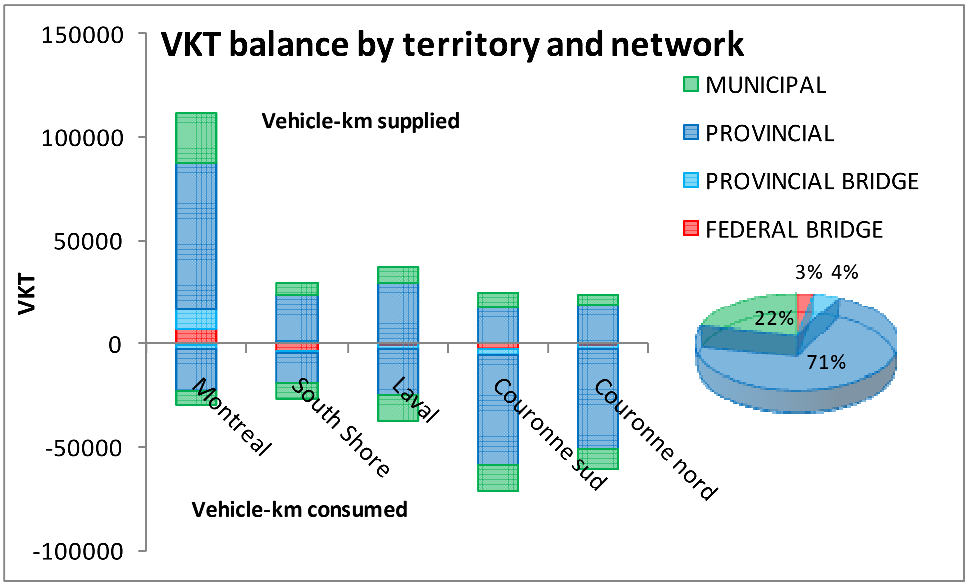


Figure 3.28: VKT consumption and supply by region and by jurisdictional network

A similar figure (not shown) can be constructed using vehicle-hours travelled but it is more meaningful to compare the time price charged by each region with the time price paid by the region's residents. This comparison is accomplished in Table 3.8 which compares the average pace of consumption with the average pace of supply. Laval and the South Shore both charge lower rates than what their residents pay. In more conventional terms, this result means that the level of service (as represented by average speed) provided by Laval and the South Shore to all users is higher than the level-of-service experienced by travellers who reside in these regions. By contrast, the two *couronnes* provide lower levels of service than those experienced by their residents on the networks of the other regions. The residents of the Island of Montreal experience the same time price as that which is supplied by their home region.

Table 3.8: Comparison of time price consumed and supplied by region

Territory of residence	PACE (Consumption – sec./km)	PACE (Supply – sec./km)
Montreal	51.5	51.4
South Shore	56.5	48.3
Laval	53.8	47.5
Couronne sud	49.3	54.9
Couronne nord	47.2	49.9
ALL	**50.6**	**50.6**

3.6 Application of simulation results to the measurement of the redistributive effects of major road infrastructure

The analysis in the previous section was based on a validation model using declared partial path information which was assumed (with some empirical support) to be reliable. This section examines the effect on an equity analysis using simulation, rather than observation, as input data. The equity indicators computed in section 3.5.5 are compared to the same statistics generated using the results of the all-or-nothing simulation model described in Chapter 3. This model successfully reproduced 74.8% of declared bridge responses. The goal is to illustrate that the simulation (predictive) model generates indicators of equity which are comparable to those derived from the validation model.

3.6.1 Trip length distributions

The initial comparison of the validation and simulation models involves a general examination of the traffic assignment results. Figure 3.29 shows the trip length distributions of the two models in terms of distance and time. Since the simulation model is an all-or-nothing assignment to the shortest path, the distribution of trip durations for the simulation model is slightly to the left of the distribution of the validation model. There is no guarantee that the simulated distribution of trip distances will also be to the left of the validated distribution, but it appears to be the case here. Table 3.9 indicates that the itineraries generated by the simulation model are on average 2.7% shorter in terms of distance and 7.7% shorter in terms of time, relative to the validation model. These two statistics represent an average speed difference of 3.9 km/h.

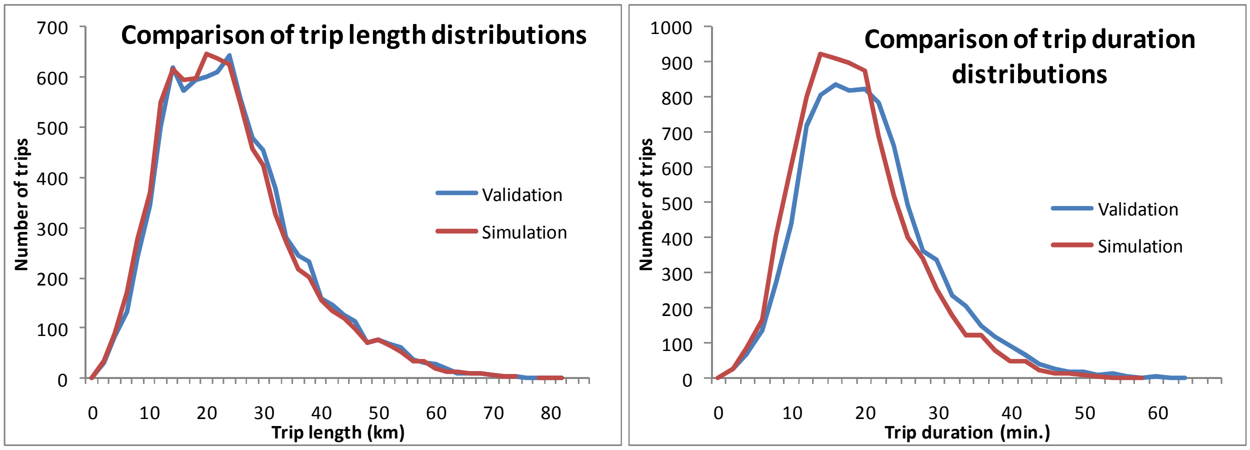


Figure 3.29: At left, the distribution of trip lengths (distances) and at right, the distribution of durations (travel times) for validated and simulated bridge responses

Table 3.9: Summary comparison of the validation and simulation models

	Validated	Simulated	Difference
VKT	218908	213042	-2.7%
Avg. length	25.5	24.8	
VHT	3076	2838	-7.7%
Avg. duration	21.5	19.8	
Avg. speed	71.2	75.1	5.5%

3.6.2 Consumption and supply

The equity analysis in the previous section was based on a comparison of the amount of transport consumed and supplied by each territory in the Greater Montreal Area. It is therefore important to verify that the simulation model does not predict drastically different quantities of supply and consumption at the territorial level. Figure 3.30 compares the amount of transport consumed by the 100 municipal sectors as calculated by the two models. Consumption is measured both in terms of distance (VKT) and time (vehicle-minutes-travelled or VmT). Apart from the systematic underestimation of consumption by the simulation model, the correspondence is almost exact.

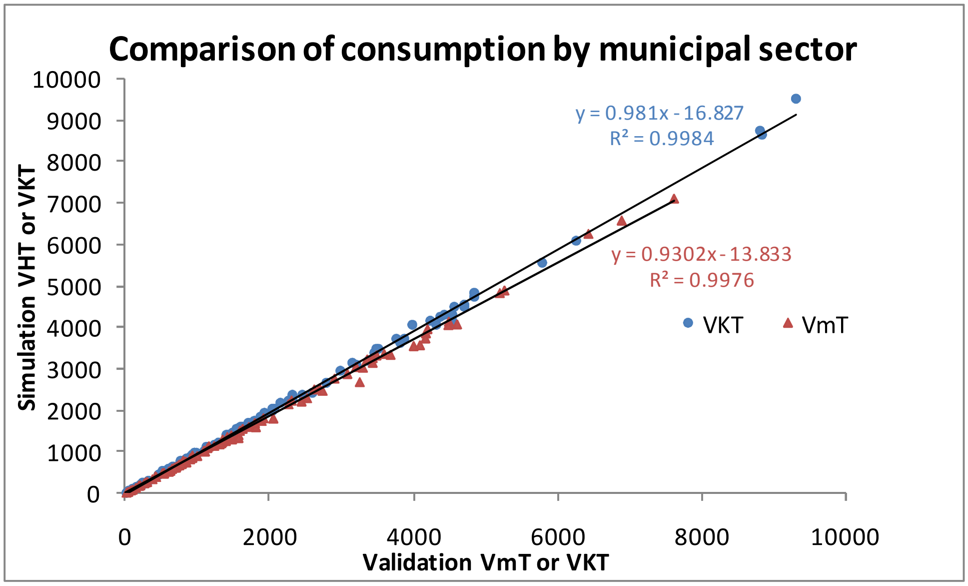


Figure 3.30: Comparison of the amount of transport consumed by municipal sector between the simulation and validation models

Figure 3.31 is a comparison of transport supplied and the results of the two models are very close to being identical, although the calculation of transport supplied appears to be slightly more sensitive to the parameters of the simulation of transport consumed but the discrepancies between the two models are very small.

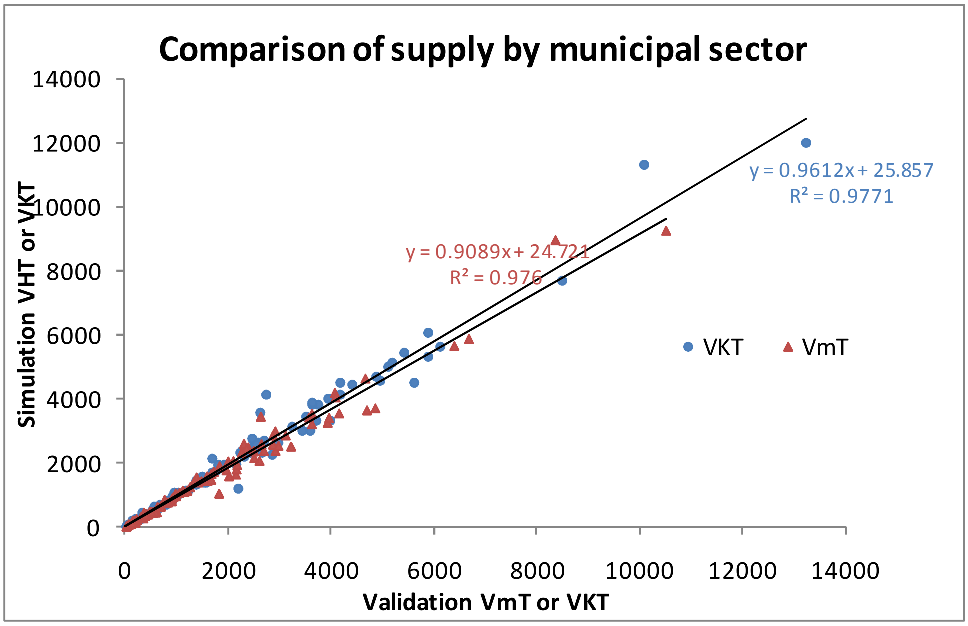


Figure 3.31: Comparison of amount of transport supplied by municipal sector between the simulation and validation models

3.6.3 The VKT balance

The influence of the simulation on model on the supply-consumption balance is tested in Figure 3.32 and Figure 3.33. Figure 3.32 plots, for the 100 municipal sectors, the supply-consumption balance predicted by the validation model versus that predicted by the simulation model. Whether the employed metric is based on VKT or time VmT, the results are almost the same. There is a very strong linear relationship between the validation and simulation model results. Based on the slope coefficients of the two regression equations, the simulation model tends to underestimate the differences between municipal consumption and supply, meaning that the inequities are slightly attenuated in the simulation model. These findings are further confirmed by Figure 3.33 which shows the VKT balance for each of the five regions of the Greater Montreal Area. The differences in results generated by the two models are almost imperceptible.

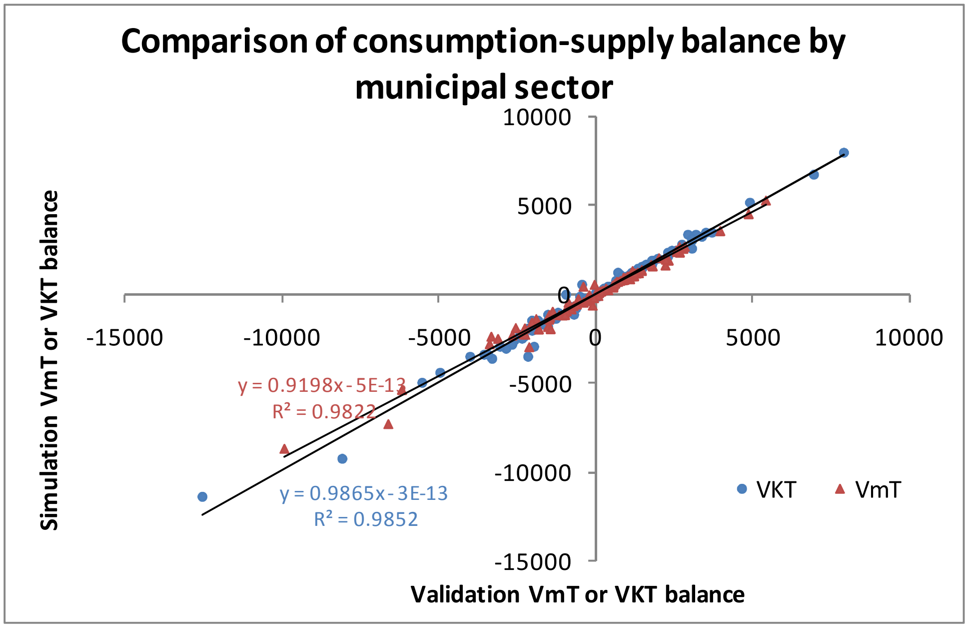


Figure 3.32: Comparison of the supply-consumption balance for the 101 municipal sectors between the simulation and validation models

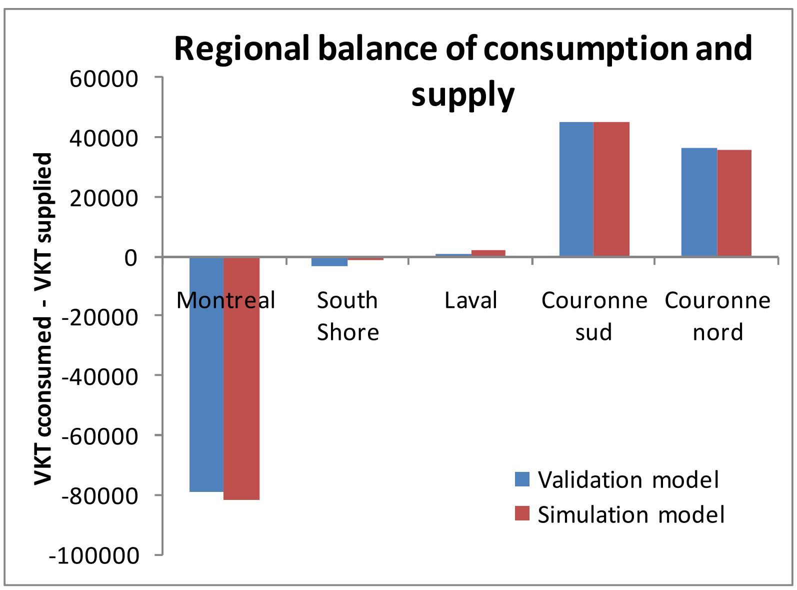


Figure 3.33: Comparison of VKT balance for the five large regions

Finally, the impact of the simulation model on the distribution of transportation resources via the various jurisdictional and functional networks is examined. Table 3.10 compares the distribution of VKT and VHT over these networks in the simulation and validation models. The discrepancies apparent in this table are much larger than those discerned in the territorial analysis above. The percentage in each cell represents a portion of the total quantity for the model in question. These totals are shown in the bottom line of the table.

Since all but two of the functional road types attract only small shares of the total traffic, the most significant differences are found in the usage patterns of freeways and arterial roads. The simulation model underestimates the use of the former and overestimates the use of the latter at both the provincial and municipal levels. This finding suggests that the attractiveness of freeways is not due solely to their time-saving attributes. Some drivers evidently prefer them over arterial roads even if their total travel time is longer.

Table 3.10: Network usage patterns based on the simulation and validation models

	VALIDATION		SIMULATION		% DIFFERENCE	
NETWORK	**VKT**	**VHT**	**VKT**	**VHT**	**VKT**	**VHT**
FEDERAL BRIDGE	**3.2%**	**3.8%**	**3.3%**	**3.5%**	**2.17%**	**-9.38%**
PROVINCIAL BRIDGE	*4.7%*	*3.8%*	*4.9%*	*3.8%*	*4.20%*	*-0.69%*
FREEWAY	*49.5%*	*38.0%*	*51.9%*	*39.0%*	*4.83%*	*2.49%*
RAMP	*5.4%*	*7.7%*	*5.1%*	*7.6%*	*-6.74%*	*-1.62%*
ARTERIAL	*13.8%*	*16.3%*	*12.4%*	*15.5%*	*-10.14%*	*-5.35%*
COLLECTOR	*0.3%*	*0.4%*	*0.3%*	*0.4%*	*-13.13%*	*-8.36%*
PROVINCIAL TOTAL	**73.8%**	**66.3%**	**74.6%**	**66.2%**	**1.07%**	**-0.17%**
RAMP	*0.5%*	*0.7%*	*0.5%*	*0.7%*	*-8.54%*	*-3.52%*
ARTERIAL	*16.9%*	*20.0%*	*16.2%*	*20.2%*	*-4.05%*	*1.22%*
COLLECTOR	*2.5%*	*3.6%*	*2.5%*	*3.8%*	*0.25%*	*5.76%*
LOCAL	*3.1%*	*5.6%*	*3.0%*	*5.6%*	*-4.36%*	*0.88%*
MUNICIPAL TOTAL	**23.0%**	**29.9%**	**22.2%**	**30.3%**	**-3.72%**	**1.58%**
ALL	**218908**	**3076**	**213042**	**2838**	**-2.68%**	**-7.75%**

3.7 Traffic models and the costs of congestion

In the context of a discussion on the pricing and financing of urban road transport supported by disaggregate, information-based analysis tools, it is important to consider alternative perspectives. This thesis has examined the problem of road pricing using a concept of equity defined in geopolitical terms and has partially assessed geopolitical equity based on directly observed usage of important infrastructure elements embodied by the major bridges of Montreal. The distribution of external transport costs associated with the use of the major bridges was compared to the distribution of the associated benefits. The existence of gaps between costs and benefits, measured at the level of politically-defined territories, illustrated the pertinence of a regional equalization mechanism for the shared financing of major road infrastructure.

This approach to the analysis of transportation-related costs and benefits is particular to the transportation planning culture within the Greater Montreal Area. The more conventional approach – covered extensively in academic and popular literature – is concerned with imputing a monetary value to the external costs of automobile travel for the purposes of establishing a road pricing regime that would charge drivers a toll to cover these external costs. This pricing system, from a certain perspective, can also be considered equitable since it requires each driver to pay the true cost of his or her consumption. It does not, however, account for the external costs of transport already borne by drivers, particularly when they are not driving.

But this conventional approach to road pricing becomes more questionable when it considers traffic congestion to be the most important external cost of associated with automobile use. This premise is problematic and is therefore worth examining in detail. Moreover, the conventional approach has been applied to the Montreal case using the same travel survey data that formed the basis for the analysis contained in this thesis. It has yet to be demonstrated that survey data alone are sufficient for the legitimate assessment of congestion phenomena. This suggestion was already made in Chapter 2, but the point should be re-emphasized if traffic congestion is recognized as a monetarily-quantifiable external cost of private automobile transport.

The evaluation of the costs of congestion in Greater Montreal has been accomplished using the concept of a congestion threshold (Gourvil & Joubert, 2004; Conseillers ADEC, 2009). According to the methodology applied in these reports, the congestion threshold is defined as an average link speed corresponding to 60% of the link free-flow speed. A link is designated as "congested" when its average speed falls below the congestion threshold. The idea of a congestion threshold is based on the fluid-flow model of traffic (Figure 1.2) where two distinct flow regimes can be identified: free-flow and forced flow. The two regimes are separated by the link capacity. The free-flow regime exists where traffic density is less than the density observed at capacity and the forced-flow regime exists where traffic density exceeds the density at capacity. Empirical data are cited to show that the congestion threshold speed corresponds to a maximum rate of traffic flow on the link or, in other words, the link capacity.

For the purposes of estimating the costs of congestion, traffic congestion is represented by a particular conception of delay that is unrelated to the intuitive notion of arriving late. Total delay (d) is calculated as the difference between the total travel cost on the link and the hypothetical total travel cost that would be incurred if the same traffic volume could use the link at the speed corresponding to the congestion threshold. In other words:

$$d = \max\left[\left(T(V) - T(C)\right)V, 0\right] \qquad (3.1)$$

where V is the total demand for the link in vehicles, $T(V)$ is the time required to cross the link when the demand is V and $T(C)$ is the time required to cross the link when the average speed on the link is 60% of the free-flow speed. This latter condition corresponds to a total demand equal to the link capacity, C. An example of the methodology is found on page 15 of (Gourvil & Joubert, 2004). Link travel times are not directly observed but are simulated using volume-delay functions similar to those discussed in section 1.3.2.4.3.

The quantity d is multiplied by an assumed monetary value of time in order to express the cost of congestion in terms of dollars. Distinct values of time are imputed to each trip purpose. Using the data in the 2003 travel survey and the calculation procedure just outlined, the total costs of congestion in the Greater Montreal Area were assessed at $1.4 billion (Conseillers ADEC, 2009). In order to illustrate that this method of calculating congestion costs is, at the very least, questionable, the following sub-section describes a short thought experiment using one of the 15 major bridges of Montreal The concept of congestion as an external cost with a theoretical monetary value is also debatable. Some points to consider in this regard are offered in the second sub-section.

3.7.1 Models of delay: experiment on a single link

The Charles-de-Gaulle Bridge in east Montreal serves as an example of a single congested link. The assertion that it is congested on a typical weekday morning is based on anecdotal evidence. Some supporting empirical evidence might be obtained by examining traffic counts collected by the Québec Ministry of Transport (Figure 3.34). The figure shows a traffic flow rate that varies considerably over the course of the day. It is notably stable around its maximum value during the morning peak period. This maximum value of approximately 6,600 veh/h is a good candidate value for the capacity of the bridge. But according to the model of congestion costs discussed above, congestion occurs when the average speed on the link falls below 60% of the free-flow speed. The calculation of congestion costs therefore requires traffic speed as input. To date, measured traffic speeds have not been used. Instead, simulation models have been constructed to estimate them. What follows is an examination of this simulation modelling framework and the implications for its use.

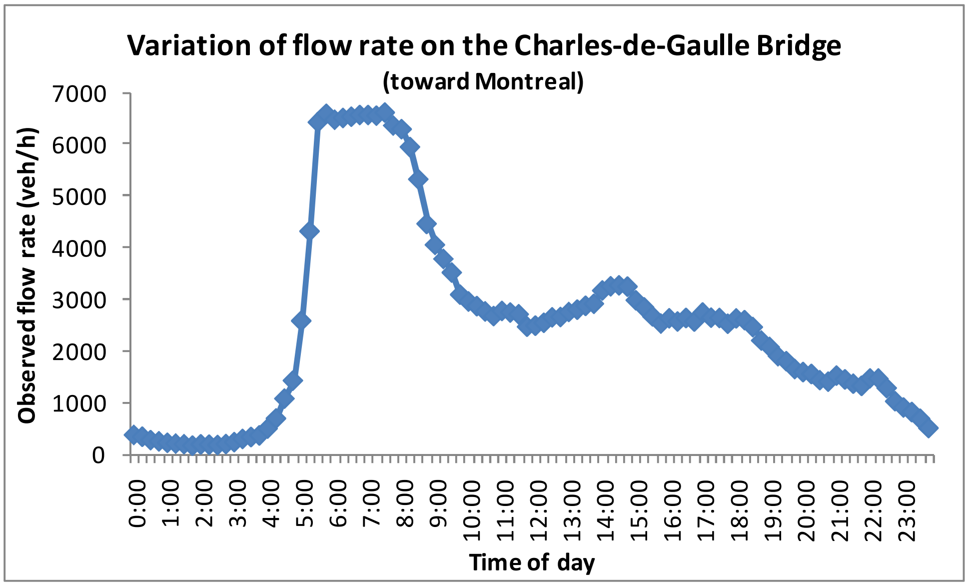


Figure 3.34: Variation of traffic toward on the Charles de Gaulle Bridge on an average weekday based on counts by the Québec Ministry of Transport

In order to simulate traffic on the bridge, it is first necessary to determine the bridge's dimensions. This bridge has a length (l) of 2 km and carries 3 freeway lanes in each direction. Only the Montreal-bound direction is analysed in this experiment. The posted speed limit on the bridge is 100 km/h and this value is used as the free-flow speed (S). The free-flow travel time of 1.2 minutes is computed as the bridge length divided by the free-flow speed. The roadside counts indicate that the maximum observed flow rate of vehicles on the bridge in the direction of Montreal is 6,600 veh/h. This rate is taken as the capacity of bridge (C). These dimensions are summarized in Table 3.11. The bridge is then simulated using two hypotheses concerning travel demand. The first hypothesis is that demand is uniform over the simulation time period. The second hypothesis is that demand varies over the course of the simulation.

Table 3.11: Attributes of the Charles-de-Gaulle Bridge

Variables	**Symbols**	**Values in the static model**
Length	l	2 km
Free flow speed	S	100 km/h
Free flow travel time	T_0	1.2 min
Number of lanes	n	3
Maximum flow rate	C or C_s	6,600 veh/h
Study period		1 h
Minimum vehicle spacing	s	10 m
Effective vehicle length	L	0.003333 km
Queue length at the link exit	Q	
Maximum queue size (physical capacity)	C_p	600 veh

3.7.1.1 The costs of congestion under uniform demand

Following the prescribed method for measuring the costs of congestion, the link is analysed using a volume-delay function and a congestion threshold. This approach is analogous to the one used in Gourvil & Joubert (2004) and Conseillers ADEC (2009). Three different hourly volumes (V) are considered in the experiment: 6,000 veh, 6,600 veh and 7,200 veh. In all three cases, the demand is assumed to be uniformly distributed over a one-hour period. This assumption is built into the volume delay function which is an expression of average cost as a function of demand over the analysis period. The volume-delay function chosen for the present experiment is the BPR function with the following form:

$$T(V) = T_0 \left[1 + \alpha \left(\frac{V}{C}\right)^{\beta}\right] \qquad (3.2)$$

where T is the travel time on the link when the volume is V, T_0 is the free-flow travel time on the link, C is the link capacity, and α and β are calibration parameters. In practice, α and β are specified in such a way that the simulated link volumes and speeds correspond as well as possible with the real link volumes and speeds at locations where such data are available. According to the literature, common values for α and β are 0.15 and 4, respectively. The α parameter is especially important in the present discussion because it defines the average speed of traffic when the link volume is equal to the link capacity. In principle, to remain coherent with the notion of a congestion threshold and the fluid flow model of traffic, α for the Charles-de-Gaulle bridge should be fixed at a value that gives a link speed equal to 60% of the free-flow speed at when volume is equal to capacity. When the link travel time at capacity is the link length divided by the speed at capacity (expressed as 0.6*S*), the parameter $\boldsymbol{\alpha}$ is uniquely determined as follows. When volume equals capacity, the ratio *V*/*C* is equal to 1. Using equation (3.2):

$$T(C) = T_0[1 + \alpha] \equiv \frac{l}{0.6S} \qquad (3.3)$$

$$\alpha = \frac{l}{0.6ST_0} - 1 \qquad (3.4)$$

But ST_0 is equal to *l*. Therefore

$$\alpha = \frac{1}{0.6} - 1 = 0.6666 \qquad (3.5)$$

In other words, if the definition of a congestion threshold of 60% of the free-flow speed is adopted, then α must be equal to 0.667 for our example link. The β parameter determines how rapidly the link travel time will increase with increasing volume. A typical value of 4 is adopted here.

By combining equations (3.1) and (3.2), the formula for total delay, d, when $V \geq C$ becomes:

$$d = \left(T_0\left[1 + \alpha\left(\frac{V}{C}\right)^{\beta}\right] - T_0[1 + \alpha]\right)V \qquad (3.6)$$

which simplifies to:

$$d = \left[\left(\frac{V}{C}\right)^{\beta} - 1\right]\alpha T_0 V \qquad (3.7)$$

This formula is used to compute delay under the three demand scenarios. A summary of particular link characteristics for each level of demand is found in Table 3.12.

Table 3.12: Indicators of link performance using a volume-delay function

	Units	Demand (V)		
Indicators	veh/hr	6000	6600	7200
Free-flow travel time	min	1.2	1.2	1.2
Free-flow total cost*	veh-hr	120.0	132.0	144.0
Average speed	km/h	68.5	60	52
Travel time	min	1.75	2.00	2.33
Total travel cost	veh-hr	174.6	220.0	280.0
Travel time at capacity	min	2.00	2.00	2.00
Total cost at capacity*	veh-hr	200.0	220.0	240.0
Average delay per vehicle	min	0.00	0.00	0.33
Total delay (Congestion cost)	veh-hr	**0.0**	**0.0**	**40.0**

* Hypothetical

The table shows, first of all, that the delay incurred when the demand is 7,200 veh is quite small. On average, it is 0.33 minutes (20 seconds) per vehicle and the total for all vehicles is 40 vehicle-hours or 14.3% of the total travel cost. Note that a demand of 7,200 vehicles corresponds to a volume-to-capacity ratio of 1.2 and average speed of 52 km/h. Moreover, when the demand is 6,000 veh, the average speed on the link is 68.5 km/h – considerably lower than free-flow speed. This reduced speed results in zero delay, as defined in equation (3.1). Finally, it must be noted that when demand is equal to capacity, the total delay is still zero. This model of congestion appears therefore to imply that no congestion costs are generated by the Charles-de-Gaulle Bridge since the level of demand never exceeds the bridge capacity. This finding illustrates a contradiction inherent to the use of volume-delay functions: the link can carry a volume that exceeds its capacity. In the case presented here, the link volume must exceed capacity if any congestion costs are to be generated.

The volume-delay function is not the only way to model congestion. An alternative (and perhaps more coherent) approach is offered by the queuing model discussed in section 1.3.1.2. All the parameters of the problem are unchanged except for α and β, which are no longer required, and the addition of a spacing parameter, s, representing the minimum distance between the front ends of two consecutive vehicles. For demonstration purposes, a reasonable value of 10 m is assumed for s. In the queuing model, the spacing parameter is used to calculate an effective vehicle length (L). For a single lane, the effective vehicle length is equal to the minimum spacing. With three lanes, however, vehicles can travel parallel to each other. As a result 10 linear metres of road can hold 3 vehicles. The effective vehicle length for the link is therefore 3.33 m in this case. These experimental parameters are listed in Table 3.11.

One of the peculiarities of the static model is that no maximum value is imposed either on the volume of traffic that can use a link or on the amount of delay that a link may generate. By contrast, the queuing model imposes two limits on the number of vehicles that can use the link: a temporal capacity and a spatial capacity. Here we define the *temporal capacity* (C_t) as the maximum hourly volume which can *exit* the link. In the present example it is set at 6,600 veh/h based on the maximum flow rate observed in Figure 3.34. In other words, a queue will form if the rate of arrivals at the link exit exceeds 6,600 veh/h. In the absence of a queue, the average speed on the link is equal to the free-flow speed. The travel time on the link becomes an increasing function of demand only when a queue starts to form at the exit. In such situations, the travel time (T_m) for a single vehicle arriving at the bridge entrance at time t is:

$$T_m(t) = \frac{Q(t)}{C_t} + \frac{l - Q(t)L}{S} \qquad (3.8)$$

where $Q(t)$ is the number of vehicles in the queue at time t. The right-hand side of this equation has two components which are added together. The first component represents the time an individual vehicle spends in the queue. This component offers a definition of delay that is less arbitrary than the one provided by a congestion threshold. It takes a value of zero as long as the arriving flow rate is lower than the temporal capacity.

The second component represents the time spent on the link while not in the queue. This component exists only if the queue length is less than the link length. This constraint on the length of the queue could be called the *spatial capacity* (C_s) of the link. The spatial capacity is determined by the amount of two-dimensional space (area) provided by the road segment. Every segment has a width expressed in terms of number of lanes n, and a length l. The division of this space by the minimum vehicle spacing (s) gives the physical capacity.

$$C_s = \frac{nl}{s} \qquad (3.9)$$

The present link has three lanes and is 2 km long and the minimum vehicle spacing is 10 m. The spatial capacity of the bridge is therefore 600 vehicles.

In order to compare the static and queuing models, it is necessary that the travel time function in the queuing model be expressed in terms of the hourly demand (V) since no time-dependent relationships are defined in the static model. To do this, the queuing phenomenon must be expressed as an average cost over the time period from t_0 to t_1:

$$T(V) = \frac{\int_{t_0}^{t_1} \frac{Q(t)}{C_t} + \frac{l - Q(t)L}{S} dt}{t_1 - t_0} \qquad (3.10)$$

If t_0 is 0 and t_1 is 1 (in units of hours) then the average is taken over the time period during which vehicles arrive at the link entrance. The above expression therefore becomes

$$T(V) = \int_0^1 \frac{Q(t)}{C_t} + \frac{l - Q(t)L}{S} dt \qquad (3.11)$$

Nothing further can be accomplished unless the evolution of the queue over time is specified. Initially, a uniform distribution of demand is assumed. In other words, vehicles enter the link at a constant rate for exactly 1 hour after which the demand falls instantaneously to 0. This assumption is implied in the static model. At any moment t, therefore, the length of the queue (Q) is the strictly positive difference between the cumulative arrivals $A(t)$ and the cumulative departures $D(t)$ at the link. In other words:

$$Q(t) = \max[0, A(t) - D(t)] \qquad (3.12)$$

The cumulative arrivals depends uniquely on the total demand and the cumulative departures depends uniquely on the link capacity.

$$A(t) = Vt \qquad (3.13)$$

$$d(t) = \begin{cases} C_t \text{ if } Q(t^-) > 0 \\ a(t) \text{ otherwise} \end{cases} \qquad (3.14)$$

where t^- is the instant just prior to t. Since

$$D(t) = \int_0^t d(u)du \tag{3.15}$$

And since the arrival rate exceeds the temporal capacity for the entire hour, then

$$Q(t) = Vt - C_t t$$

$$Q(t) = (V - C_t)t \tag{3.16}$$

Using equations (3.11) and (3.16) the expression for average travel time in the presence of a queue becomes

$$T(V) = \int_0^1 \frac{t(V - C_t)}{C_t} + \frac{l - tL(V - C_t)}{S} dt \tag{3.17}$$

$$T(V) = \frac{t^2(V - C_t)}{2C_t} + \frac{lt}{S} - \frac{t^2 L(V - C_t)}{2S}\Bigg|_0^1 \tag{3.18}$$

$$T(V) = \frac{(V - C_t)}{2}\left(\frac{1}{C_t} - \frac{L}{S}\right) + \frac{l}{S} \tag{3.19}$$

Using equations (3.1) and (3.19), the expression for total delay using the queuing model when $V \geq C_t$ is:

$$d = \left[\frac{(V - C_t)}{2}\left(\frac{1}{C_t} - \frac{L}{S}\right) + \frac{l}{S} - \frac{l}{S}\right]V$$

$$d = \left[\frac{(V - C_t)}{2}\left(\frac{1}{C_t} - \frac{L}{S}\right)\right]V \tag{3.20}$$

This formula generates the link performance characteristics shown in Table 3.13.

Table 3.13: Indicators of link performance using a queuing model

	Units	Demand (*V*)		
Indicators	veh/h	6000	6600	7200
Free-flow travel time	min	1.2	1.2	1.2
Free-flow total cost	veh-h	120	132	144
Avg. speed	km/h	100	100	36.0
Avg. Travel time	min	1.2	1.2	3.3
Total travel cost	veh-h	120	132	399.3
Avg. Travel time at capacity	min	1.2	1.2	1.2
Total cost at capacity	veh-h	120	132	144
Average delay per vehicle	min	0	0	2.13
Total delay (Congestion cost)	veh-h	0	0	**255.3**
Max queue length	veh	0	0	600
	m	0	0	2000
Max travel time	min	1.2	1.2	5.5

Using the queuing model, the total delay is much greater than that obtained using the volume-delay function (255 veh-h vs. 40 veh-h). Moreover, the delay calculated using the queuing model represents 64% of the total travel cost (255 veh-h out of 399 veh-h) whereas using the volume-delay function, the delay represents only 14.3% of the total travel cost (40 veh-h out of 280 veh-h). Finally, the demand of 7,200 vehicles generates a maximum queue length equal to the length of the bridge and is therefore the highest uniform demand that can be coherently simulated without adding an upstream link.

In order for the volume-delay function to generate a delay comparable to that calculated using the queuing model, either the incoming volume or the β parameter must be increased significantly. The effects of doing both are illustrated in Figure 3.35. Box d1 represents the delay generated by setting the β parameter equal to 15. Box d2 represents the delay resulting from a total demand of 8,800 vehicles. The total delay in both cases is equal to 255 veh-h – the same value obtained using the queuing model. The legitimacy of using 15 as a value for β is questionable. It should also be noted that a total demand of 8,800 vehicles would produce a queue 5.3 km longer than the link itself. In summary, the volume-delay model is difficult to reconcile with the observed supply and demand characteristics of the Charles-de-Gaulle Bridge.

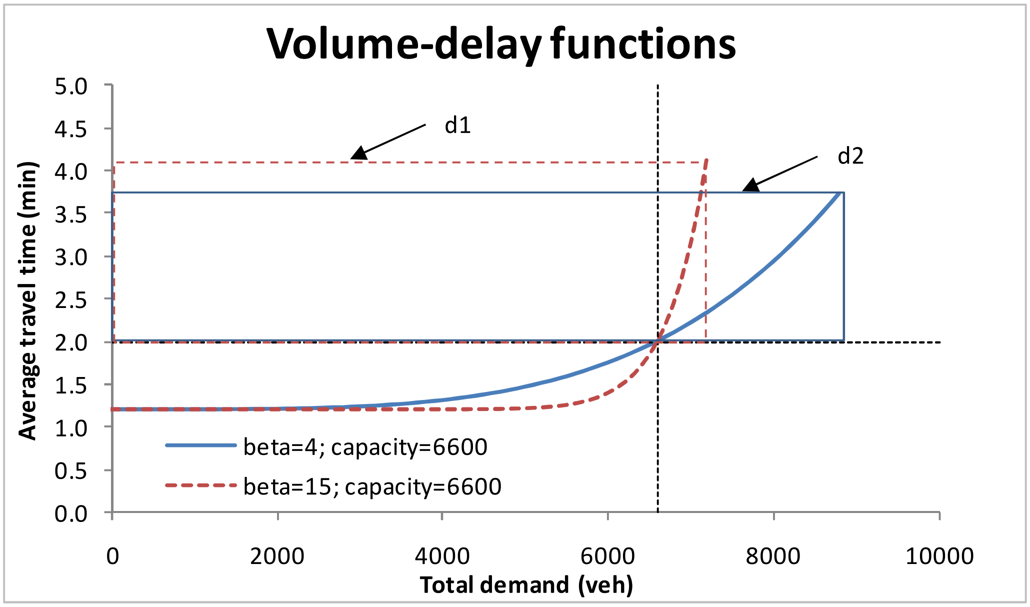


Figure 3.35: Total delay generated by two different volume-delay functions

3.7.1.2 The costs of congestion under non-uniform demand

While the volume-delay model requires the assumption of a uniform distribution of demand, the queuing model does not. The queuing model can be used to illustrate the importance of the demand distribution in the estimation of delay caused by traffic congestion. A commonly-used distribution in queuing systems is the Erlang distribution whose probability density function is given by:

$$f(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \qquad (3.21)$$

A convenient characteristic of the Erlang distribution is that its shape can be modified using the parameters k and λ. Regardless of the exact form of the distribution of arrivals, any plausible non-uniform distribution implies the existence of a peak level of demand. This peak will be considerably higher than the average demand for the period. For example, under uniform demand, it was shown that 7,200 vehicles was the largest traffic volume that would generate a queue shorter than the bridge itself. Under a non-uniform distribution, however, the queue can exceed the bridge length even if the total demand is considerably lower than 7,200 vehicles. This effect is illustrated in Figure 3.36 which is based on an Erlang distribution of arrivals with both *k* and *λ* equal to 5. The figure shows a queue that begins to form after about 15 minutes and is reabsorbed after nearly 45 minutes. Even though the queue is of brief duration, it reaches the link entrance (2 km upstream from the exit) after about 15 minutes. Also, despite generating a very long queue, the total demand of 4,150 vehicles is much less than the temporal capacity of the bridge. If the volume-delay model were applied in this circumstance, no delay would be calculated on the bridge.

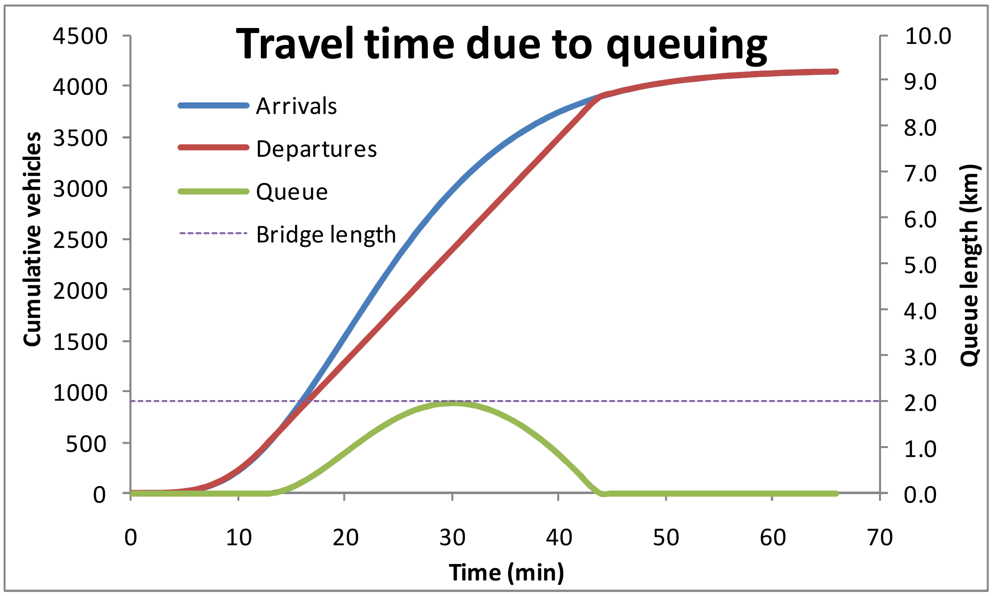


Figure 3.36: Queuing model using a non-uniform demand distribution

3.7.1.3 Summary Conclusions

This experiment has demonstrated the inconsistencies implied by the use of a static traffic assignment model for the calculation of congestion costs. First, the definition of congestion imposes a restriction on one of the calibration parameters of the volume delay function. Second, the volume-delay function simulates significant congestion effects only at physically impossible levels of demand. Third, the amount of delay on a link depends not only on the amount of incoming traffic, but also on its temporal distribution – an element which cannot be incorporated into the static model.

The queuing model generates a much higher estimate of delay than the volume-delay model. The enormous discrepancies between the results of the two models, as well as the numerous arbitrary elements of the volume-delay function and its associated congestion threshold, call into question the credibility of congestion cost estimates based on simulated congestion. Direct observation of traffic speeds and flow rates are required for a legitimate assessment of congestion costs, although the concept itself should be scrutinized further.

3.7.2 Congestion costs, equilibrium and the implications for equitable transport

While the use of static models to estimate congestion costs presents some methodological difficulties, the practice of classifying recurrent traffic congestion as a cost on society is problematic from the perspective of equitable transportation. Aside from the already-argued proposition that the operating objectives of the transport system extend beyond mere efficiency, the soundness of quantifying congestion costs is questionable on the grounds that recurrent congestion represents a suboptimal equilibrium between supply and demand (Stopher, 2004a). The traditional notion of congestion costs has developed from the economist's habit of equating inefficiency with loss. In addition, if one accepts that a user-equilibrium represents a suboptimal state then the cost associated with inefficiency can only be calculated by comparing the user-equilibrium with the most efficient optimum: the system-equilibrium. The difficulty associated with such an analysis is that it implies the assumption of much greater responsibilities by road system operators. In order to approach a system-optimum condition, the road network must be equipped, at the very least, with a control system that measures traffic conditions in real time and with a credible means of relaying relevant information to vehicles. Such systems are costly to implement and operate. It is much easier for a road operator to charge its users a toll for the use of the road based on the hypothesis that the users themselves are responsible for the inefficient allocation of resources represented by prevailing traffic conditions. This hypothesis appears to be the basis for imputing a monetary cost to recurrent congestion.

Moreover, the notion that time spent in congested traffic represents an economic loss is thrown into question by the observed behaviour of drivers. It has previously been proposed that the monetary value of travel time experienced by private citizens in the course of their day-to-day activities is impossible to quantify (see section 3.2.1). Rather than attempt to attach a dollar amount to time spent travelling, it may be more fruitful to recall the concept of equilibrium. In principle, a generalized equilibrium includes not only the transportation system but also the activity patterns of individuals. Within strict time constraints, people plan their daily routine in such a way that they are able to sleep, eat, work, care for their families and indulge in discretionary pleasures. The effectiveness of the scheduling effort depends on each person's awareness of the duration and the scheduling flexibility of a given activity. For most activities, these parameters are easily determined. Travel is one activity where the amount of time required is not known with certainty, at least initially. With experience, travellers construct an expectation (an average) of the time needed to complete a particular journey and plan their daily schedule accordingly. The equilibrium state of the transportation network in general, resulting from the choices of all travellers with respect to departure time, mode and route, arises from this expectation. Under stable circumstances, an individual has no incentive to change his travel patterns since it will require him to reschedule other activities.

If, on the other hand, the equilibrium is disrupted by an extreme event such as the complete closure of a freeway or a major snowstorm, the personal scheduling system fails. People arrive late, miss appointments, are forced to cancel activities and so forth. The monetary costs associated with these events are theoretically measurable and of significant magnitude, yet little effort is made to quantify them. From this perspective, the costs of congestion are generated not by the bad habits of drivers, but by failures of the road transport system. Inadequate maintenance will require road closures and unsafe design induces costly traffic accidents. By ignoring the costs of non-recurrent congestion, the method applied to measure the costs of congestion in Greater Montreal allows road network operators to evade accountability for the proper functioning of their own systems.

3.8 Conclusions

This chapter has demonstrated a methodology for assessing the degree to which the current operation of major road facilities, as represented by observed travel behaviour, constitutes equitable transport policy. Using information in the travel survey, it was possible to ascertain the distribution among different population groups of the external costs of road transport associated with automobile travel on the major bridges. This assessment was performed using a validation model that correctly reproduces 100% of the observed bridge choices. However, it was shown that the use of an assignment model which correctly reproduces nearly 75% of observed facility choices yields aggregate indicators of equity which do not differ significantly from those generated by validation model.

With specific reference to the Montreal case, some general conclusions can be drawn. First, the primary beneficiaries of the major bridges are residents of the suburban communities surrounding the Island of Montreal. Second, the Island of Montreal assumes nearly half the total cost of transportation related to the use of the major bridges. Third, the suburban regions closest to Montreal (Laval and the South Shore) are equitably treated by the fifteen major bridges in the sense that the benefit they obtain is almost equal to the costs they incur. The equality of costs and benefits in these regions is due to the important volumes of through-traffic using the network on their territory. The outlying *couronnes* extract the greatest benefit and are subject to the lowest external cost burden. These regions are currently experiencing the highest rate of population growth within the Greater Montreal Area.

These findings do not imply that some of the major bridges of Montreal should be closed or even that they should be tolled. Both these measures amount to an increase of transportation costs within the region and, while they would certainly alter travel patterns, they would do little to address the distribution problems outlined above. In fact, bridge users living in Laval, the South Shore and in Montreal would be penalized by these measures even though current traffic patterns suggest that they are not net beneficiaries of bridge infrastructure when external costs are considered. Rather, these findings are further evidence in support of the theory that major road facilities linking the central city to suburban communities tend to encourage the development of the suburban communities at the direct expense of the central city. Tolls are instruments too blunt to rectify the underlying cost-benefit imbalance.

Furthermore, this study has illustrated the importance of the provincial level of government in the distribution of road transport resources. Although the provincial network is estimated to account for 12.8% of the entire Greater Montreal road system (see Table 2.3), it supplies 74% of regional VKT and 34% of regional VHT associated with the use of the 15 major bridges (11 of which are provincially controlled). This domination of the road transport market is attributable to the low time price of travel (high speeds) offered by provincial infrastructure. Although all citizens of Quebec pay for the construction and maintenance of these facilities, the benefits are not necessarily distributed evenly throughout the population. The major bridges examined in the present research project provide a clear example of cost-benefit imbalances between different population groups.

A possible method for establishing a more equitable regional transportation system is to make municipalities pay for the infrastructure their citizens *use*, rather than just the infrastructure that lies within their jurisdiction. In the Greater Montreal Area, for example, the regional transport budget for major road infrastructure could be financed using contributions from each municipality, and these contributions would be proportional to the amount of the vehicle-kilometres consumed by residents. At the same time, the external costs generated by drivers using major road infrastructure can be imputed to particular territories based on the location of driver households. The external cost burden carried by each municipality can be included in the calculation of each municipality's contribution. In this way, citizens already paying the external costs of road transport would be charged a reduced amount for their own consumption. This financing scheme constitutes an equalization mechanism which, in principle, would allow for the realization of fiscal parity among the municipalities that make up the Greater Montreal Area. A similar mechanism is already in place for the financing of major public transport infrastructure. Although quantifying consumption and supply of road transport at the municipal level represents a methodological challenge, this research has demonstrated that such a calculation is feasible. Moreover, it is feasible through a partial representation of travel demand derived from a travel survey that samples 4.5% of the region's households and a totally disaggregate approach to traffic simulation. This approach is less intrusive than the continuous tracking of individual vehicles using GPS and surveillance technology that would be required for obtaining exact distributions of the costs and benefits of automobile travel.

Clearly, major bridges and freeways are not the only types of public infrastructure which play a role in the distribution of transportation costs and benefits. High-speed high-capacity mass transit networks have similar effects on the structure of travel demand. While the equity effects of the two systems analysed separately have been documented in the present study and elsewhere (Chapleau, 1995; Chapleau & Morency, 2005), the supply-demand interaction between the public transit and public road system is worth a brief comment.

The users of public transit derive direct benefits from their use of the public transit system. But they also confer a benefit on the users of a congested road network by not consuming space on the road. It is not immediately clear that this benefit is reciprocal. While automobile users pay taxes dedicated to financing the public transit system, most of the subsidy goes to maintaining services outside the principal corridors and time periods of peak demand. The heavy transit infrastructure is largely self-financing at times when the road network is at its most congested.

The suggestion that the users of public transit are delivering an uncompensated benefit to auto drivers presents an instructive example of the equitable transportation problem. Many people, sometimes by choice but sometimes by force of circumstance, must endure the significant hardships associated with public transit use. These hardships vary between localities, but in Montreal they include walking and waiting outside for extended periods in a harsh climate, unscheduled delays due to equipment failures and security threats, and extremely crowded conditions in vehicles and stations. As a direct result of this willing or imposed self-sacrifice, the users of the road system, free and independent in their climate-controlled vehicles, are able to complete their journey considerably faster.

CHAPTER 4 CONCLUSION

4.1 Summary of the work

This thesis has presented a methodology for the application of the totally-disaggregate approach to the simulation of road infrastructure and the evaluation of equitable road transport. A sample of the 2003 Montreal travel survey was isolated and validated. The sample consisted of 8,583 trips containing declarations describing the use of a single major bridge during a typical morning peak period. This sample of observed demand for the major bridges of Montreal was assigned to a complete model road network comprising over 100,000 links and 70,000 nodes. A totally disaggregate assignment of trips to the network generated complete itineraries for each trip in the survey subsample.

Two types of model were constructed: a validation model and a simulation model. The validation model assigned each trip to the shortest time path incorporating the declared bridge. This model was used to compare observed and simulated travel patterns and to compute indicators of consumption. The simulation models adopted a variety of approaches to assign trips to the network. Probabilistic and deterministic methods of assignment were used. The probabilistic models analysed driver bridge choice using an analogy with public transport whereby each bridge is treated as a “line” with particular service attributes. The deterministic models used the traffic simulation tools provided in the open-source TRANSIMS package. All the simulation models were capable of reproducing around 75% of bridge declarations contained in the travel survey. The itineraries generated by the validation model were used to compute indicators of road transport consumption by households as well as indicators of road transport supply by each geopolitical entity in the region. This exercise illustrated the existence of a cost-supply imbalance that favours the off-island municipalities at the expense of the city of Montreal. The quantification of this imbalance at the level of sub-regions or municipalities was then compared to a more conventional approach to the calculation of external road transport costs embodied by recent studies on the monetary costs of traffic congestion in the Greater Montreal Area.

4.2 Original contributions

This thesis makes several original contributions to the domain of large-scale traffic simulation and to the body of work concerning equitable transportation and infrastructure financing. First, this research describes a method for the totally disaggregate analysis of model errors using a confusion matrix. This type of error analysis permits the identification biases in the model and can reveal ways in which they can be corrected. This thesis also contains a detailed discussion on the indifference of drivers presented with almost equivalent alternative options. Indifference is one of the main causes of model prediction errors and is distinguished from two other types of prediction error: deviance and gross error. Third, this thesis represents the first application of real (non-synthesized) totally disaggregate travel demand data in TRANSIMS, a state-of-the art disaggregate travel behaviour simulation environment.

The analysis of equitable transport which is made possible by the totally disaggregate approach to transportation planning is augmented through the classification of network links by jurisdiction (municipal, provincial or federal). The incorporation of jurisdiction into the model revealed the important role of the provincial government, through its provision of high-speed high-capacity road infrastructure, in the inequitable redistribution of transport costs and benefits within a large urban region. Also, the same analysis formed the basis for an innovative road pricing system applicable to important infrastructure elements (exemplified by the major bridges of Montreal) that accounts for the distribution of external road transport costs among the users and non-users of the road network. The proposed pricing system constitutes an equalization mechanism for achieving parity between the numerous jurisdictions coexisting within a single urban area. The distributions of road transport costs and benefits among the population of the Greater Montreal Area were initially estimated using a model that reproduces 100% of the observed bridge choices. When a simulation model that reproduces 75% of observed bridge choices is used instead, only a small impact on the aggregate measures of benefits and costs is observed. This experiment demonstrates the applicability of a predictive model to an assessment of transport costs and benefits associated with specific road infrastructure elements at the sub-regional level. Finally, this thesis has demonstrated that an approximately equitable road pricing mechanism can be established using surveys of a sample of the travelling population without having to resort to sophisticated methods for tracking private vehicles as they move through the road network.

The concept of geopolitical equity and its evaluation using disaggregate assignment methods did not originate in this thesis. This thesis has, however, provided an in-depth examination of infrastructure elements that clearly illustrate the problems arising from the redistributive effects of public transportation infrastructure. The major bridges of Montreal constitute the primary system for transport and exchange of people, vehicles and goods between different political jurisdictions within the Greater Montreal Area. At the municipal level, the quest for fiscal parity between governments having equivalent responsibilities often leads to disputes when superior levels of government (the province or the federation) use collective resources for the benefit of a select few. These disputes complicate immensely all manner of discussions concerning urban sprawl, greenhouse gases, road tolls and the sustainability of transport systems. If the superior level of government must intervene in the provision of transportation within an urban region, it must also provide an equalization mechanism judged by all parties to be fair and legitimate. It is for this reason that a section of Chapter 3 is devoted to dissecting estimates of the “costs of congestion”. The methods used to propagate the message that current road usage patterns in Greater Montreal are inefficient have been, to date, based on arbitrary thresholds and simplifying assumption disguised with complicated algebra. The claims that are made based upon these types of analysis must be scrutinized and evaluated, especially if equitable road transport is to become an achievable goal. This thesis has provided a small amount of the necessary scrutiny in addition to representing, through its detailed study of the major bridges of Montreal, an incremental advance toward the development of an equitable and transparent infrastructure financing regime.

4.3 Limitations of the research approach

The present research was limited by several factors. First, in order to demonstrate the feasibility of the two disaggregate simulation methods represented by the multinomial logit model and TRANSIMS, the sample of travel demand was restricted to data from the morning peak period of the 2003 Montreal survey. The use of a sub-sample of a single travel survey means that the sample size is relatively small (less than 9,000 trip observations) compared to the number of observations that could be obtained through the fusion of the four travel surveys that have collected bridge declarations. Although the smaller sample size made the simulation process faster and more manageable, additional facets of driver behaviour could be examined if multiple surveys were combined and if travel over the 24 hours of an average day was included in the analysis. A larger sample might also permit the analysis of periodic bridge closures, accidents and weather events. Such studies constitute essential topics for future research.

A second important limitation is the treatment of congestion effects. Although attempts were made to represent congestion effects using both probabilistic and deterministic assignment models, the representation was incomplete. While some sort of volume to capacity ratio can be constructed for each of the fifteen bridges, no attempt was made to reproduce congestion elsewhere in the network. This omission is not due to a lack of analysis tools or methods but rather to a glaring absence of important information. In the Greater Montreal Area, most of the existing knowledge of road congestion is anecdotal. Average traffic speeds and volumes are measured only punctually – at single locations and at particular moments in time. Data describing the real-time measurement of queues and other delay phenomena are practically non-existent. Equally inconvenient from both the strategic and operational planning perspectives is the absence of a searchable database of road infrastructure components. Relevant information that would ideally be found in such a database include, for each street segment, the number of lanes, the posted speed limit, the parking regime, the presence of a bus line, the control system (signalling and signage), a measured capacity and a measured free-flow speed. All these elements would permit the construction of an improved model of traffic congestion.

Third, the use of unexpanded survey records in the calculation of road transport consumption and supply is a potential obstacle to the application of the equalization scheme developed in Chapter 3. Although the demographic expansion factors provide a plausible estimate of the magnitude of numerous travel phenomena, they are difficult to apply in a disaggregate traffic simulation (see section 2.5.2.1).

4.4 Research Perspectives

The limitations just described suggest numerous directions for future work. First, with respect to the representation of road transport supply, effort should be devoted to the detailed and realistic codification of the Montreal regional road network. Link attributes such as functional class, speed and capacity were determined for the 15 major bridges in the present project but would ideally be coded for freeways and urban streets throughout the metropolitan area. The estimation of capacity and free-flow traffic speeds would be based on direct measurement and physical characteristics of the built infrastructure such as the number of lanes and lane geometries. Detailed information on the control system would also be required. The measurement of traffic flows and speeds over extended periods of time at multiple locations is essential for acquiring an improved understanding of road traffic phenomena and would require a proper instrumentation of the superior road network (bridges, freeways and arterial roads).

Secondly, with respect to the representation of road transport demand, the estimates of road consumption and supply should be recomputed using weighted survey observations (expansion factors). The challenge is to find a method of expanding survey records that yields vehicle volumes closer to those observed on the road network since realistic representations of vehicle demand are necessary to properly model congestion effects. The expansion of trip records is especially difficult in when performing microsimulation since each individual vehicle and its physical attributes must be represented. The incorporation of commercial vehicles is another important component of this objective.

A third potential area of exploration involves the probabilistic approach to facility choice. The facility choice models developed in this thesis included attributes of the facility and associated path. Attributes of the driver were not found to be significant predictors of bridge choice but additional experiments involving a larger sample of trips might be more fruitful.

The enlargement of the sample of observed bridge choices constitutes a fourth research perspective. The question concerning the choice of bridge has been included in four consecutive travel surveys conducted in 1993, 1998, 2003 and 2008. The fusion of data from these four surveys would facilitate the study of more detailed aspects of traveller behaviour. Obviously, the sample size could be expanded further by examining not only the morning peak period but the entire 24 hours of a typical weekday. Such an analysis should include an investigation of trip chains and the role of the trip chain structure in the choice of route and road facility.

In addition, the detailed analysis of itineraries demonstrated in section 2.5.2.5 should be generalized for the entire sample of observed trips. A comparison of the attributes of the simulated and validated itineraries would likely provide further insight into driver behaviour. Such an analysis would be especially meaningful if performed using a validated model road network and if additional information on the non-bridge portions of each itinerary was known. An example of such information is the declarations of freeway use which have been included in multiple travel surveys. While it appears that the validation of freeway declarations would be considerably more complicated than the validation of bridge declarations because of the greater complexity of the freeway network and the often ambiguous naming conventions applied to freeways in the Greater Montreal Area, these issues, are not insurmountable and efforts should be made to address them with the goal of improving the legitimacy of traffic assignment models.

Finally, the use of certain historical records in conjunction with the survey data would permit the analysis of extreme events such as major road closures and adverse weather conditions. The examination of these unpredictable occurrences would reveal how drivers respond to short-term disruptions to the network equilibrium.

For many centuries, civil engineers have been responsible for the construction of major public works. It is only recently that civil engineers have been asked to take some responsibility for the impacts of these public works on the public itself. At the moment, the prevailing cultural trend seems focused on the costs of providing a universally accessible public road network. Traffic congestion, pollution and energy consumption are seen as threats to the collective well-being. These concerns have, perhaps temporarily, overshadowed the tremendous benefits of automobile travel which has produced historically unprecedented levels of personal mobility and flexibility. While the debate over whether the benefits exceed the costs continues, the engineer remains preoccupied with the optimal functioning of the system for which he is responsible. Economic inequities and distortions are at the heart of consumption-related problems but they must be clearly identified and understood before they can be resolved.

REFERENCES

Achim, C., & Florian, M. (1979). *EMME: a computer system for transportation planners* (No. 129). Université de Montréal: Centre de recherche sur les transports.

Allard, B., & Grondines, J. (2007). *Algorithme d'arrimage des relevés GPS provenant des minibus de Transport Adapté de la STM avec les données SIG-Transport.* Paper presented at Plan d'intervention, St-Hyacinthe, QC.

Alonso, W. (1964). *Location and land use: toward a general theory of land rent*. Cambridge, MA: Harvard University Press.

Añez, J., de la Barra, T., & Pérez, B. (1996). Dual graph representations of transport networks. *Transportation Research Part B, 30*(3), 209-216.

Arnott, R., de Palma, A., & Lindsey, R. (1993). A structural model of peak-period congestion: a traffic bottleneck with elastic demand. *American Economic Review, 83*(1), 161-179.

Arnott, R., & Small, K. A. (1994). The Economics of Traffic Congestion. *American Scientist, 82*, 446-455.

Atkins, S. T. (1977). Transportation planning: is there a road ahead? *Traffic Engineering & Control, 18*(2), 58-62.

Atkins, S. T. (1984). Why Value Travel Time? The Case Against. *Highways and Transportation, 31*(7), 2-7.

Atkins, S. T. (1986). Transportation planning models - what the papers say. *Traffic Engineering & Control, 27*(9), 460-467.

Baass, K. (2003). *Les variables fondamentales* [Lecture notes] Département des genies civils, géologiques et des mines, École Polytechnique de Montréal, Canada.

Bar-Gera, H. (1999). *Origin-based algorithms for Transportation Network Modeling* (No. 103). Research Triangle Park, NC: National Institute of Statistical Sciences.

Barceló, J., & Casas, J. (2006). Stochastic Heuristic Dynamic Assignment Based on AIMSUN Microscopic Traffic Simulator. *Transportation Research Record, 1964*, 70-80.

Barth, D. (2009) The bright side of sitting in traffic: Crowdsourcing road congestion data. Retrieved October 11, 2009, from http://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html

Beckmann, M., McGuire, C. B., & Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Santa Monica, California: The Rand Corporation.

Bekhor, S., & Reznikova, L. (2007). *Application of the Cross Nested Logit Route Choice Model in Stochastic User Equilibrium Traffic Assignment*. Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Bellman, R. (1958). On a routing problem. *Quarterly of Applied Mathematics, 16*, 87-90.

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H., & Mishalani, R. (1998). *DynaMIT: a simulation-based system for traffic prediction*. Paper presented at the DACCORD Short Term Forecasting Workshop, Delft, The Netherlands.

Bergeron, D. (1991). *Approche désagrégée appliquée à l'analyse des réseaux de transport urbain.* M.Sc.A. École Polytechnique de Montreal, QC., Canada.

Bergeron, D., & Chapleau, R. (1996). *Modélisation de l'accessibilité à l'île de Montréal depuis l'enquête-ménage Origine-Destination STCUM-MTQ de 1993.* Paper presented at the 31st annual meeting of the l'Association québécoise du transport et des routes, Québec, QC.

Bierlaire, M., & Frejinger, E. (2008). Route choice modelling with network-free data. *Transportation Research Part C, 16*, 187-198.

Bohte, W., & Kees, M. (2008). *Deriving and validating trip destinations and modes for multi-day GPS based travel surveys: An application in the Netherlands*. Paper presented at the 87th Annual Meeting of the Transportation Research Board, Washington, DC.

Bonsall, P., & Kelly, C. (2005). Road user charging and social exclusion: The impact of congestion charges on at-risk groups. *Transport Policy, 12*, 405-418.

Braess, D., Nagurney, A., & Walkobinger, T. (2005). On a Paradox of Traffic Planning. *Transportation Science, 39*(4), 446-450.

Buliung, R., & Kanaroglou, P. S. (2006). A GIS toolkit for exploring geographies of household activity/travel behavior. *Journal of Transport Geography, 14*, 35-81.

Bureau of Public Roads. (1964). *Traffic Assignment Manual*. Washington, D.C.: U.S. Department of Commerce, Urban Planning Division.

Cascetta, E., Russo, F., Viola, F. A., & Vitetta, A. (2002). A model of route perception in urban road networks. *Transportation Research Part B, 36*, 577-592.

Chapleau, R. (1974). *Réseaux de transport en commun: structure informatique et affectation.* Ph.D., École Polytechnique de Montréal, Qc., Canada.

Chapleau, R., Allard, B., & Canova, M. (1982). *MADITUC, un modèle de planification opérationnelle adapté aux entreprises de transport en commun de taille moyenne* (No. 265). Montréal: Centre de recherche sur les transports de l'Univérsité de Montréal.

Chapleau, R. (1988). *Mesure de la redistribution des bénéfices et des coûts associés à un réseau de transport en commun.* Paper presented at the 23rd annual meeting of the Association Québecoise des transports et des routes.

Chapleau, R. (1992). *La modélisation de la demande de transport urbain avec une approche totalement désagrégée*. Selected Proceedings of the 6th World Conference on Transport Research, Lyon, France, (Vol. 2, pp. 937-948).

Chapleau, R. (1995). Measuring Impacts of Transit Financing Policy in Geopolitical Context: Montreal Case. *Transportation Research Record, 1496*, 52-58.

Chapleau, R. (1997). *Conducting Telephone Origin-Destination Household Survey with an Integrated Informational Approach*. Transport Surveys: Raising the Standard – Proceedings of an International Conference on Transport Survey Quality and Innovation, Grainau, Germany.

Chapleau, R., & Chu, K. K. A. (2007). *Modeling transit travel patterns from location-stamped smart card data using a disaggregate approach*. Paper presented at the 11th World Conference on Transportation Research, Berkeley, California.

Chapleau, R., & Morency, C. (2004). *Effets redistributifs des infrastructures de transport routier et en commun du Grand Montréal, 1987-1998*. Paper presented at the Dix-septièmes entretiens du centre Jacques-Cartier, Lyon, France.

Chapleau, R., & Morency, C. (2005). *Le transport urbain équitable: cas de la Grande Région de Montréal*. Paper presented at the 40th annual meeting of the Association québécoise du transport et des routes, Laval, Québec.

Chapleau, R., & Piché, D. (2009). *AGIRAS-TU: Approche Globale Intégrée de Recherche et d'Analyse Systémique du Transport Urbain*. Paper presented at the 44th annual meeting of the Association québécoise du transport et des routes, Montréal, Québec.

Cheah, L., Dalton, P., & Hariri, O. (1992, 1992). *Alternative Approaches to Volume-Delay-Function Formulation.* Paper presented at the 7th International EMME/2 Users' Conference, Vancouver, British Columbia.

Chou, J., Takriti, S., & Underwood, S. (1993). *Behavioral Modeling of Network Traffic: Integrating Route Choice and Traffic Assignment*. Ann Arbor, Michigan: University of Michigan.

Chu, K. K. A., & Chapleau, R. (2008). Enriching archived smart card transaction data for transit demand modelling. *Transportation Research Record, 2063*, 63-72.

Conseillers ADEC Inc. (2009). *Évaluation des coûts de la congestion routière dans la région de Montréal pour les conditions de référence de 2003*. Montreal.

Crouse, D. L., Goldberg, M. S., & Ross, N. A. (2009). A prediction-based approach to modelling temporal and spatial variability of traffic-related air pollution in Montreal, Canada. *Atmospheric Environment, 43*, 5075-5084.

Crouse, D. L., Ross, N. A., & Goldberg, M. S. (2009). Double burden of deprivation and high concentrations of ambient air pollution at the neighbourhood scale in Montreal, Canada. *Social Science & Medicine, 69*, 971-981.

Crozet, Y., & Marlot, G. (2001). Péage urbain et ville "soutenable": figures de la tarification et avatars de la raison économique. *Cahiers Scientifiques du transport, 40*, 79-113.

Dial, R. B. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research, 5*, 83-111.

Dial, R. B. (1976). Urban Transportation Planning System: Philosophy and Function. *Transportation Research Record, 619*, 43-38.

Dial, R. B. (2000). Minimal-revenue congestion pricing Part II: An efficient algorithm for the general case. *Transportation Research Part B, 34*, 645-665.

Dial, R. B. (2006). A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B, 40*, 917-936.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik, 1*, 269-271.

Downs, O. B., Chapman, C. H., Cahn, R. C., & Hersch, J. S. (2006). *Assessing road traffic speed using data obtained from mobile data sources*. United States Patent No. 20070208501 A1. U. S. P. a. T. Office. Retrieved October 14, 2009 from www.google.ca/patents.

Essakali, M. D. (1999). *Expérimentation à propos des effets redistributifs de l'usage d'un réseau routier urbain.* M.Sc.A. École Polytechnique de Montréal, Québec, Canada.

Federal Highway Administration. (2005). *TRANSIMS Fundamentals*.

Florian, M., Chapleau, R., Nguyen, S., Achim, C., James-Lefebvre, L., Galarneau, S., et al. (1979). Validation and Application of EMME: An Equilibrium Based Two-Mode Urban Transportation Planning Method, *Transportation Research Record*, 728, 14-22.

Florian, M., Constantin, I., & Florian, D. (2009). *A new look at the projected gradient method for equilibrium assignment*. Paper presented at the 88th Annual Meeting of the Transportation Research Board, Washington, DC.

Ford, L. R., & Fulkerson, D. R. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, 399-404.

Forkenbrock, D. J., & Sheeley, J. (2004). *Effective Methods for Environmental Justice Assessment* (National Cooperative Highway Research Program Report 532). Washington DC: Transportation Research Board.

Frank, M., & Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly, 3*, 95-110.

Frejinger, E. (2008). *Route choice analysis: data, models, algorithms and applications.* Ph.D., École Polytechnique Fédéral de Lausanne, Lausanne, Switzerland.

Gartner, N. H., & Wagner, P. (2004). Analysis of Traffic Flow Characteristics on Signalized Arterials, *Transportation Research Record,* 1883, 94-100. Washington, DC: Transportation Research Board.

Gilliéron, P.-Y., & Waegli, A. (2005). Galileo et le prix de la mobilité. *Route et traffic*, 3, 19-22.

Gourvil, L., & Joubert, F. (2004). *Évaluation de la congestion routière dans la région de Montréal* (No. RTQ-04-01): Ministère des transports de Québec.

Habib, K. M. N. (2007). *Modelling activity generation processes.* Ph.D., University of Toronto, Canada.

Hassin, R., & Haviv, M. (2003). *To Queue or Not to Queue: Equilibrium Behavior in Queuing Systems*. Boston: Kluwer Academic Publishers.

Hess, S., Bierlaire, M., & Polak, J. (2005). Capturing taste heterogeneity and correlation structure with Mixed GEV models. In R. Scarpa & A. Alberini (Eds.), *Applications of Simulation Methods in Environmental and Resource Economics* (pp. 55-76). Dordrecht, The Netherlands: Springer.

Hess, S., Rose, J. M., & Polak, J. (2008). *Non-trading, lexicographic and inconsistent behaviour in SP choice data*. Paper presented at the 87th Annual Meeting of the Transportation Research Board, Washington, DC.

Hitchcock, F. L. (1941). Distribution of a product from several sources to numerous localities. *Journal of Mathematics and Physics, 20*, 224-230.

Holden, D. J. (1989). Wardrop's Third Principle. *Journal of Transport Economics and Policy, 23*(3), 239-262.

Hoogendoorn-Lanser, S., & Bovy, P. (2007). *Modeling overlap in multi-modal route choice by inclusion of trip part specific path size factors*. Paper presented at the TRB 86th Annual Meeting, Washington DC.

Horowitz, A. J. (1991). *Delay-Volume Relations for Travel Forecasting: Based on the 1985 Highway Capacity Manual* (FHWA-PD-92-015). Federal Highway Administration: Washington, DC.

Horowitz, A. J. (2001). Computational Issues in Increasing the Spatial Precision of Traffic Assignments. *Transportation Research Record, 1777*, 68-74.

Horowitz, A. J. (2004). *Reference Manual Quick Response System II Version 7*. [Software]. Milwaukee, WI: AJH Associates.

Hu, M. B., Jiang, R., Wu, Y.-H., Wang, W.-X., & Wu, Q.-S. (2008). Urban traffic from the perspective of the dual graph. *European Physical Journal B, 63*, 127-133.

Institute of Transportation Engineers. (2008). *Trip generation handbook* (8^{th} Edition). Washington, D.C.

Jan, O., Horowitz, A. J., & Peng, Z. R. (2000). Using GPS Data to Understand Variations in Path Choice. *Transportation Research Record, 1706*, 37-44.

Jayakrishnan, R., Tsai, W. T., Prashker, J. N., & Rajadhyaksha, S. (1994). *A faster path-based algorithm for traffic assignment.* Paper presented at the 73rd Annual Meeting of the Transportation Research Board, Washington, D.C.

Koenigsberg, E. (1966). On jockeying in queues. *Management Science, 12*(5), 412-436.

Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning, 30*, 271-274.

Lakshmanan, T. R., Nijkamp, P., Rietveld, P., & Verhoef, E. T. (2001). Benefits and costs of transport. *Papers in Regional Science, 80*(2), 139-164.

Lämmer, S., Gehlsen, B., & Helbing, D. (2006). Scaling Laws in the Spatial Structure of Urban Road Networks. *Physica A, 363*(1), 80-95.

Lari, A., & Iacono, M. (2006). *Transportation finance, congestion, and equity: some policy perspectives*. Paper presented at the 85th Annual Meeting of the Transportation Research Board, Washington, DC.

Lee, C., & Machemehl, R. B. (2005). *Combined Traffic Signal Control and Traffic Assignment: Algorithms, Implementation and Numerical Results*. Austin, Texas: University of Texas at Austin.

Levinson, D. M. (2002). Identifying Winners and Losers in Transportation. *Transportation Research Record, 1812*, 179-185.

Litman, T. (2006). *Evaluating Transportation Equity.* Victoria Transport Policy Institute.

Mahmassani, H. S. (2001). Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications. *Networks and Spatial Economics, 1*, 267-292.

Mahmassani, H. S. (2006). Integrating Activity-based Demand and Network Assignment Models.

Mahut, M., Florian, M., Florian, D., Velan, S., & Tremblay, N. (2005). *Equilibrium dynamic traffic assignment for large, congested networks*. Montreal: INRO.

Manheim, M. L. (1979). *Fundamentals of Transportation Systems Analysis Volume 1: Basic Concepts*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Martin, W. A., & McGuckin, N. A. (1998). *Travel estimation techniques for urban planning* (National Cooperative Highway Research Program Report 365). Washington, DC: Transportation Research Board.

Maruyama, T., & Sumalee, A. (2007). Efficiency and equity comparison of cordon and area-based road pricing schemes using a trip chain equilibrium model. *Transportation Research Part A, 41, 655-671.*

May, A. (1990). *Traffic flow fundamentals*. Englewood Cliffs, New Jersey: Prentice Hall.

May, A., & Milne, D. S. (2000). Effects of alternative road pricing systems on network performance. *Transportation Research Part A, 24*, 407-436.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics, 3*, 303-328.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*, 447-470.

Meneguzzer, C. (1995). An equilibrium route choice model with explicit treatment of the effect of intersections. *Transportation Research Part B, 29*(5), 329-356.

Miller, E., Roorda, M. J., & Carrasco, J. A. (2005). A tour-based model of travel mode choice. *Transportation, 32*, 399-422.

Miller, H. J. (1999). Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T). *Geographical Analysis, 31*, 373-399.

Mohring, H. (2006). Congested roads: an economic analysis with Twin Cities' illustrations. In G. Roth (Ed.), *Street Smart* (pp. 141-170). New Brunswick, NJ: Transaction Publishers.

Morency, C. (2004). *Contributions à la modélisation totalement désagrégée des interactions entre mobilité urbaine et dynamiques spatiales.* Ph.D. École Polytechnique de Montréal, Québec, Canada.

Nazer, Z., & Pickford, A. (2007). *Advanced technologies for road user charging.* Paper presented at the The 86th Annual Meeting of the Transportation Research Board, Washington, D.C.

Orfeuil, J.-P. (2008). *Une approche laïque de la mobilité*. Paris: Descartes et cie.

Ortuzar, J. D. (1983). Nested logit models for mixed-mode travel in urban corridors. *Transportation Research A, 17A*(4), 283-299.

Papinski, D., Scott, D. M., & Doherty, S. (2009). Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS. *Transportation Research Part F, 12*, 347-358.

PB Consult. Inc. (2002). *Réalisation des études d'achalandages et renevus pour les projets autoroutiers et partenariat public-privé dans la région de Montréal*. Ministère des transports du Québec.

Peeta, S., & Ziliaskopoulos, A. K. (2001). Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future. *Networks and Spatial Economics,* 1, 233-265.

Porta, S., Crucitti, P., & Latora, V. (2006). The network analysis of urban streets: A dual approach. *Physica A, 369,* 853-866.

Ramming, M. S. (2002). *Network knowledge and route choice.* Ph.D. Massachusetts Institute of Technology, Cambridge, Massachusetts.

Scarlett, M. (1970). *The Travel Behaviour of Urban Commuters.* Ph.D. University of Pittsburgh, Pennsylvania.

Serras, J. (2007). *Multidimensional Multilevel Representation for Traffic Simulation Models.* Ph.D. The Open University, United Kingdom.

Sheffi, Y. (1985). *Urban Transportation Networks*. Englewood Cliffs, New Jersey: Prentice-Hall.

Simon, H. A. (1957). *Models of Man: Social and Rational.* New York: John Wiley and Sons, Inc.

Simon, H. A. (1996). *The Sciences of the Artificial* (3rd Edition). Cambridge, Massachusetts: MIT Press.

Speiss, H. (1997). *Conical Volume-Delay Functions*. Aegerten, Switzerland: EMME/2 Support Center.

Spurr, T. (2005). *Construction of a Detailed Deterministic User-Equilibrium Traffic Assignment Model for the Greater Montreal Area Using Geographic Information Systems*. M.Eng. McGill University, Montreal, Canada.

Spurr, T., & Chapleau, R. (2007). *The contribution of the confusion matrix to the analysis of mode choice in Montreal*. Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Statistics Canada. (2003). *Census of Canada 2001*.

Stopher, P. R. (2004a). Reducing road congestion: a reality check. *Transport Policy, 11*, 117-131.

Stopher, P. R. (2004b). Spatial data issues: a historical perspective. In D. A. Hensher, K. J. Button, K. E. Haynes & P. R. Stopher (Eds.), *Handbook of Transport Geography and Spatial Systems*: Elsevier.

Stopher, P. R., FitzGerald, C., & Xu, M. (2007). Assessing the accuracy of the Sydney Household Travel Survey with GPS. *Transportation, 34*(6), 723-741.

Stopher, P. R., & Stecher, C. (Eds.). (2006). *Travel Survey Methods: Quality and Future Directions*. Oxford: Elsevier.

Talvitie, A. P. (1997). Things planners believe in, and things they deny. *Transportation, 24*, 1-31.

Talvitie, A. P. (2007). *Theoryless Planning*. Paper presented at the 11th World Conference on Transport Research, Berkeley, CA.

Tétrault, P., & el-Geneidy, A. (2009). *Estimating bus run times for new limited stop service using archived AVL and APC data*. Paper presented at the 88th Annual Meeting of the Transportation Research Board, Washington, DC.

Tobler, W. R. (1988). The quadratic transportation problem as a model of spatial interaction patterns. In *Geographical Systems and Systems of Geography* (pp. 75-88). London, Ontario: Department of Geography of the University of Western Ontario.

Trannoy, A. (2007). Équité territoriale, acceptabilité et grandes infrastructures de transport. In Le calcul économique (pp. 402-431). Paris: Economica.

Transport Canada. (2006). *Le coût de la congestion urbaine au Canada*. Ottawa.

Transport Canada. (2009). *Transportation in Canada 2008: An overview*. Ottawa.

Transportation Research Board. (2000). *Highway Capacity Manual*. Washington, DC.

Transportation Research Board. (2007). *Metropolitan Travel Forecasting*. Washington, DC.

Tremblay, P. (2007). *Modélisation des transports à l'échelle régionale: survol des approches utilisées*. Paper presented at the 42nd annual meeting of the Association québécoise du transports et des routes, Montreal, Québec.

Trépanier, M., Chapleau, R., & Allard, B. (2002). *Geographic information system for transportation operations: models and specificity*. Paper presented at the 4th Transportation Specialty Conference, Montreal, Québec.

U.S. Department of Transportation Federal Highway Administration. (2007). *Manual on Uniform Traffic Control Devices*. Washington, DC.

van Nes, R. (2002). *Design of multimodal transport networks a hierarchical approach*. Ph.D. Delft University, The Netherlands.

Vickrey, W. (1969). Congestion theory and transport investment. *American Economic Review, 59*, 251-261.

Volchenkov, D., & Blanchard, P. (2007). Transport Networks Revisited: Why Dual Graphs? Retrieved from http://www.arxiv.com/PS_cache/arxiv/pdf/0710/0710.5494v1.pdf.

Vovsha, P., & Bradley, M. (2006). Advanced Activity-Based Models in Context of Planning Decisions. *Transportation Research Record*, *1981*, 34-41.

Vrtic, M., Schuessler, N., Erath, A., & Axhausen, K. W. (2007). *Design elements of road pricing schemes and their acceptability*. Paper presented at the 86th Annual Meeting of the Transportation Research Board, Washington, DC.

Wardrop, J. G. (1952). Some theoretical aspects of road traffic research. *Institution of Civil Engineers*, *Proceedings Part II*, *1*, 325-378.

Webster, F. V., & Cobbe, B. M. (1966). *Traffic Signals*. London: Her Majesty's Stationery Office.

Weisbrod, G. E., Vary, D., & Treyz, G. (2001). *Economic Implications of Congestion* (National Cooperative Highway Research Program Report 463). Washington DC: Transportation Research Board.

Yang, H., & Zhang, X. (2002). Multiclass network toll design problem with social and spatial equity constraints. *Journal of transportation Engineering, 128*(5), 420-428.

Yerra, B. M., & Levinson, D. M. (2005). The emergence of hierarchy in transportation networks. *Annals of Regional Science, 39*(3), 541-553.

Zahavi, Y. (1979). *The 'UMOT' Project.* Washington, DC: USDOT; Ministry of Transport, Fed. Rep. of Germany.